

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 31, 2002, 08:58:58 ; Search time 4437.29 Seconds
(without alignments)
17388.122 Million cell updates/sec

Title: US-09-917-384-2
Perfect score: 3687
Sequence: 1 atggagcgaacccaacatc.....tcagctgcacagccagctga 3687

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenBank:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_lm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_inv:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
|------------|-------|-------------|--------|-------|-------------|

| 1 | 483.6 | 13.1 | 3240 | 1 | CFICELASE | L25809 Cellulomona |
|----|-------|------|--------|---|------------|---------------------|
| 2 | 427 | 11.6 | 3004 | 1 <th>ACU33212</th> <th>U33212 Acidothermu</th> | ACU33212 | U33212 Acidothermu |
| 3 | 427 | 11.6 | 3004 | 6 <th>I23701</th> <th>I23701 Sequence 6</th> | I23701 | I23701 Sequence 6 |
| 4 | 339.4 | 9.2 | 2535 | 1 <th>AB078006</th> <th>AB078006 Streptomy</th> | AB078006 | AB078006 Streptomy |
| 5 | 324 | 8.8 | 3404 | 6 <th>169280</th> <th>169280 Sequence 1</th> | 169280 | 169280 Sequence 1 |
| 6 | 324 | 8.8 | 3503 | 1 <th>TFU18978</th> <th>U18978 Thermomonas</th> | TFU18978 | U18978 Thermomonas |
| 7 | 320 | 8.7 | 41906 | 1 <th>SC567</th> <th>AL031515 Streptomy</th> | SC567 | AL031515 Streptomy |
| 8 | 314.6 | 8.5 | 1269 | 6 <th>169283</th> <th>169283 Sequence 4</th> | 169283 | 169283 Sequence 4 |
| 9 | 314.6 | 8.5 | 1365 | 1 <th>AF268074</th> <th>AF268074 Thermomon</th> | AF268074 | AF268074 Thermomon |
| 10 | 293 | 7.9 | 213050 | 1 <th>AL646079</th> <th>AL646079 Ralstonia</th> | AL646079 | AL646079 Ralstonia |
| 11 | 275.6 | 7.5 | 1231 | 1 <th>AF172344</th> <th>AF172344 Cellulomo</th> | AF172344 | AF172344 Cellulomo |
| 12 | 264.2 | 7.2 | 11541 | 1 <th>AE003960</th> <th>AE003960 xyella f</th> | AE003960 | AE003960 xyella f |
| 13 | 181.6 | 4.9 | 3227 | 1 <th>AF200304</th> <th>AF200304 Caldicbac</th> | AF200304 | AF200304 Caldicbac |
| 14 | 171 | 4.6 | 4367 | 1 <th>AF163837</th> <th>AF163837 Caldicbac</th> | AF163837 | AF163837 Caldicbac |
| 15 | 134.2 | 3.6 | 11707 | 1 <th>AF078737</th> <th>AF078737 Caldicell</th> | AF078737 | AF078737 Caldicell |
| 16 | 134.2 | 3.5 | 11707 | 6 <th>E35099</th> <th>E35099 Truncated c</th> | E35099 | E35099 Truncated c |
| 17 | 130.6 | 3.5 | 5437 | 1 <th>CASR692X</th> <th>AF036924 Caldicell</th> | CASR692X | AF036924 Caldicell |
| 18 | 117.2 | 3.2 | 165849 | 2 <th>AC096270</th> <th>AC096270 Rattus no</th> | AC096270 | AC096270 Rattus no |
| 19 | 116.4 | 3.2 | 2977 | 6 <th>A28170</th> <th>A28170 B. lautus st</th> | A28170 | A28170 B. lautus st |
| 20 | 114.8 | 3.1 | 2831 | 1 <th>BACCBLAA</th> <th>M76588 Bacillus la</th> | BACCBLAA | M76588 Bacillus la |
| 21 | 114.6 | 3.1 | 3262 | 1 <th>AF078038S1</th> <th>AF078038 Caldicell</th> | AF078038S1 | AF078038 Caldicell |
| 22 | 113.6 | 3.1 | 3776 | 1 <th>CFIB14C</th> <th>L38827 Cellulomona</th> | CFIB14C | L38827 Cellulomona |
| 23 | 110 | 3.0 | 131995 | 14 <th>OPU75930</th> <th>U75930 Orxyia pseu</th> | OPU75930 | U75930 Orxyia pseu |
| 24 | 108.4 | 2.9 | 749 | 14 <th>NEPORP2</th> <th>D13755 Multicapsid</th> | NEPORP2 | D13755 Multicapsid |
| 25 | 107 | 2.9 | 183685 | 2 <th>AL645600</th> <th>AL645600 Mus muscu</th> | AL645600 | AL645600 Mus muscu |
| 26 | 104 | 2.8 | 13669 | 14 <th>TFV1TP</th> <th>X14855 Thermoprote</th> | TFV1TP | X14855 Thermoprote |
| 27 | 101.2 | 2.7 | 3480 | 1 <th>CFICEND</th> <th>L02544 Cellulomona</th> | CFICEND | L02544 Cellulomona |
| 28 | 100.8 | 2.7 | 1621 | 1 <th>THEFE2AA</th> <th>M73321 Thermomonas</th> | THEFE2AA | M73321 Thermomonas |
| 29 | 100.2 | 2.7 | 4034 | 14 <th>TFV1TP</th> <th>X14717 Thermoprote</th> | TFV1TP | X14717 Thermoprote |
| 30 | 98.8 | 2.7 | 6005 | 1 <th>AF078038S5</th> <th>AF078042 Caldicell</th> | AF078038S5 | AF078042 Caldicell |
| 31 | 98.8 | 2.7 | 6416 | 6 <th>E35100</th> <th>E35100 Truncated c</th> | E35100 | E35100 Truncated c |
| 32 | 98.6 | 2.7 | 3701 | 1 <th>AF172724</th> <th>AF172724 Caulobact</th> | AF172724 | AF172724 Caulobact |
| 33 | 98.6 | 2.6 | 11283 | 1 <th>AE005710</th> <th>AE005710 Caulobact</th> | AE005710 | AE005710 Caulobact |
| 34 | 96.4 | 2.6 | 5513 | 1 <th>AT286105</th> <th>Z86105 A. thermophl</th> | AT286105 | Z86105 A. thermophl |
| 35 | 96.2 | 2.6 | 6510 | 6 <th>AX251402</th> <th>AX251402 Sequence</th> | AX251402 | AX251402 Sequence |
| 36 | 96.2 | 2.6 | 6510 | 6 <th>AX251811</th> <th>AX251811 Sequence</th> | AX251811 | AX251811 Sequence |
| 37 | 96.2 | 2.6 | 6510 | 6 <th>AX345297</th> <th>AX345297 Sequence</th> | AX345297 | AX345297 Sequence |
| 38 | 96 | 2.6 | 5439 | 1 <th>CDCCELA</th> <th>L32742 Caldocellum</th> | CDCCELA | L32742 Caldocellum |
| 39 | 95.2 | 2.6 | 1116 | 1 <th>AF233376</th> <th>AF233376 Streptomy</th> | AF233376 | AF233376 Streptomy |
| 40 | 95.2 | 2.6 | 1116 | 6 <th>AR129926</th> <th>AR129926 Sequence</th> | AR129926 | AR129926 Sequence |
| 41 | 95.2 | 2.6 | 1116 | 6 <th>AR168360</th> <th>AR168360 Sequence</th> | AR168360 | AR168360 Sequence |
| 42 | 94.8 | 2.6 | 330 | 6 <th>AR135147</th> <th>AR135147 Sequence</th> | AR135147 | AR135147 Sequence |
| 43 | 93.8 | 2.5 | 4977 | 1 <th>CDCMANA</th> <th>L01257 Caldocellum</th> | CDCMANA | L01257 Caldocellum |
| 44 | 93.4 | 2.5 | 786 | 6 <th>AX339682</th> <th>AX339682 Sequence</th> | AX339682 | AX339682 Sequence |
| 45 | 93.4 | 2.5 | 1877 | 6 <th>AX339680</th> <th>AX339680 Sequence</th> | AX339680 | AX339680 Sequence |

ALIGNMENTS

RESULT 1
LOCUS CFICELASE 3240 bp DNA linear BCT 20-JUL-1994
DEFINITION Cellulomonas fimi cellulase, complete cds.
ACCESSION I25809
VERSION I25809.1 GI:456028
KEYWORDS cellulase.
SOURCE Cellulomonas fimi (library: ATCC 484) DNA.
ORGANISM Cellulomonas fimi

REFERENCE
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Micrococcales; Cellulomonadaceae; Cellulomonas.
1 (bases 1 to 3240)
Meinke, A., Gilkes, N.R., Kwan, E., Kilburn, D.G., Warren, R.A. and
Miller, R.C., Jr.

TITLE Cellulohydrolase A (cbha) from the cellulolytic bacterium
Cellulomonas fimi is a beta-1,4-exocellulohydrolase analogous to
Tichoderma reesei CBH II
Mol. Microbiol. 12 (3), 413-422 (1994)
94344030

JOURNAL MEDLINE
FEATURES
source location/Qualifiers
1. 3240
/organism="Cellulomonas fimi"
/db_xref="taxon:1708"

Db 1957 GACACGACGCGCCGTCCTCCGCGCCG 1984

RESULT 2
ACU33212 3004 bp DNA linear BCT 19-SEP-1995
LOCUS Acidothermus cellulolyticus E I beta-1,4-endoglucanase precursor
DEFINITION gene, complete cds.
ACCESSION U33212
VERSION U33212.1 GI:988299
KEYWORDS
SOURCE
ORGANISM Acidothermus cellulolyticus.
Acidothermus cellulolyticus
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Frankineae; Acidothermaceae; Acidothermus.
REFERENCE
AUTHORS Laymon, R.A., Himmel, M.E. and Thomas, S.R.
TITLE Direct Submission
JOURNAL Submitted (04-AUG-1995) Steven R. Thomas, Applied Biological Sciences, National Renewable Energy Laboratory, 1617 Cole Blvd., Golden, CO 80401, USA

FEATURES
source location/Qualifiers
1..3004 /organism="Acidothermus cellulolyticus"
/strain="11b (ATCC 43068)"
/db_xref="taxon:28049"
stem_loop 710..725
RBS 762..769
repeat_region 781..817
/note="three direct repeats of ACGGCGCA"
/rpt_type=direct
/rpt_unit=781..788
824..2512
/codon_start=1
/transl_table=11
/product="E I beta-1,4-endoglucanase precursor"
/protein_id="AAU75477.1"
/db_xref="GI:988300"
/translation="MPRLRRVGSRMVLRGVVAVLALVALNALNLAVPRPARAG
GYWHSGREIIDANNVPRVAGINVFETCTCAVYVHGASRDYRSMLOIKSLGYNTI
RLPYSDILKRETMNSINFYMNODLOGISLQYMDKIVAYAGIGIRIILIDRRAPD
CSGSALEMTSSVSERATWISDQLAQRKRPVYEDLNEPBDPACWGGSDSID
WRLAERAGNAVLISVNPMLLFVEGVSYNGSYWGNLQAGAGYPVVLANPRLVY
SAHDYATSVYPTQTFPSDPTFPNNMPGIMNKNMGYLFNONIAPVMLGEFTGLQSTTDQ
TWLKLTVYLRPTAOYGADSFQMTFMSNNPDGGLIKDMQTVDYKDGXLAPIK
SIIIDPVGASAPSSQSPSPSPSASRTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT
ASPT
TNSMAAYTQNGQSVTARNMSYNNVIOPEQNTTFGFQASVTSNAAPVACAS"

sig_peptide 824..946
mat_peptide 947..2509
/EC_number="3.2.1.4"
/note="cellulase"
/function="cleaves beta-1,4-glycosidic linkages of cellulose"
/product="E I beta-1,4-endoglucanase"
stem_loop 2514..2560
BASE COUNT 604 a 931 c 938 g 531 t
ORIGIN

Query Match 11.6%; Score 427; DB 1; Length 3004;
Best Local Similarity 99.8%; Pred. No. 2e-32;
Matches 438; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 atggaagcgaacccaacaatccggaagcgaagctaccagaagaagacagaatg 60
DB 2567 ATGGAGCGAACCACAAATCCGACGCAACTGCAGGTACAGAGAGACGACCAATG 2626
QY 61 cccgcacatcctcaaaacgagctgagcggcgctctcgccggggcggtgagcatcgagcc 120
DB 2627 CCCGCATCTCAAAACGCGTGCAGCCGCGCTCTCGCCGGGGCGGTGACATCGCAGCC 2686

QY 121 tccatcgctccgctgagcagatgcagcatctgcatcgccgagcagcagctgcacatcc 180
DB 2687 TCCATCGTGCCTGGCGATGACGATCCCTCCATCGCCCGACGACGTCGCAATGCC 2746
QY 181 tatgcgaggagcagcctctctgcaaccgctactggcgcaaaagtacagagcgaagc 240
DB 2747 TATGCGGAGCGACCTTCTTCTGTCACCCCTACTGGCGCAAAAGTACGAGACGAA-CG 2805
QY 241 ggaacacagacaaatgcacatcgcagcagcgaatgagcgctgttccacatctgagc 300
DB 2806 GCGAACACAGACCAATGCGACTCTCGACGCGAAATATGGCGCTGTTCCATATTTGCAGC 2865
QY 301 gccgcttgatgtagcgcagcatcgctgcatacagcgctcaagcgagaccggcttgagc 360
DB 2866 GCCGCTGATGTGAGACCGCATCGCTGCATCAACGGCGTCAACGGCGGACCCGGCTTGAGC 2925
QY 361 acatatctgagcagccgcccctctccagcagcaaggaaccccccctgaagcatggatc 420
DB 2926 ACATATCTGAGACGCGCCCTCTCCAGACAGGGAACACCCCTGAAGTCATTGAGATT 2985
QY 421 gtcatacagatcgccg 439
DB 2986 GTCATCTACGATCTGCCG 3004

RESULT 3
LOCUS 123701 3004 bp DNA linear PAT 07-OCT-1996
DEFINITION Sequence 6 from patent US 5536655.
ACCESSION 123701
VERSION 123701.1 GI:1603571
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 3004)
AUTHORS Thomas, S.R., Laymon, R.A. and Himmel, M.E.
TITLE Gene coding for the E1 endoglucanase
JOURNAL Patent: US 5536655-A 6 16-JUL-1996;
FEATURES
source location/Qualifiers
1..3004 /organism="unknown"
BASE COUNT 604 a 931 c 938 g 531 t
ORIGIN

Query Match 11.6%; Score 427; DB 6; Length 3004;
Best Local Similarity 99.8%; Pred. No. 2e-32;
Matches 438; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 atggaagcgaacccaacaatccggaagcgaagctaccagaagaagacagaatg 60
DB 2567 ATGGAGCGAACCACCAATCCGACGCACTGCAGTACAGAGAGACGACCAATG 2626
QY 61 cccgcacatcctcaaaacgagctgagcggcgctctcgccggggcggtgagcatcgagcc 120
DB 2627 CCCGCATCTCAAAACGCGTGCAGCGCGGCTCTCGCCGGGGCGGTGACATCGCAGCC 2686
QY 121 tccatcgctccgctgagcagatgcagcatctgcatcgccgagcagcagcagctgcacatcc 180
DB 2687 TCCATCGTGCCTGGCGATGACGATCCCTCCATCGCCCGACGACGTCGCAATGCC 2746
QY 181 tatgcgaggagcagcctctctgcaaccgctactggcgcaaaagtacagagcgaagc 240
DB 2747 TATGCGGAGCGACCTTCTTCTGTCACCCCTACTGGCGCAAAAGTACGAGACGAA-CG 2805
QY 241 ggaacacagacaaatgcacatcgcagcagcgaatgagcgctgttccacatctgagc 300
DB 2806 GCGAACACAGACCAATGCGACTCTCGACGCGAAATATGGCGCTGTTCCATATTTGCAGC 2865
QY 301 gccgcttgatgtagcgcagcatcgctgcatacagcgctcaagcgagaccggcttgagc 360
DB 2866 GCCGCTGATGTGAGACCGCATCGCTGCATCAACGGCGTCAACGGCGGACCCGGCTTGAGC 2925

| | | | |
|------------|---|--|----------------------------|
| Oy | 361 | acatctcgagcgcgcgcctctccacgagcaggaacacccctgaaactttagact | 420 |
| Db | 2926 | ACATATCTGGACGCGCCCTCTCCACGACGACGACACCCCTGAAGTCAATTGATT | 2985 |
| Oy | 421 | gtcatctacagatctgcgag | 439 |
| Db | 2986 | GTCACTCTACGATCTGCCG | 3004 |
| RESULT | 4 | | |
| LOCUS | AB078006 | 2535 bp | DNA linear BCT 19-JAN-2002 |
| DEFINITION | Streptomyces sp. M23 gene for exoglucanase CBHI, complete cds. | | |
| ACCESSION | AB078006 | | |
| VERSION | AB078006.1 | GI:18250950 | |
| KEYWORDS | | | |
| SOURCE | | | |
| ORGANISM | Streptomyces sp. M23 (strain:M23) DNA, clone_11b:pUC 118 Streptomyces sp. M23 genomic DNA. | | |
| REFERENCE | | | |
| AUTHORS | 1 (sites) | | |
| TITLE | Park,C., Kawaguchi,T., Sumitani,J. and Arai,M. | | |
| REFERENCE | Cloning and sequencing of an exoglucanase gene from Streptomyces sp. M23, and its expression of Streptomyces lividans | | |
| AUTHORS | Unpublished | | |
| TITLE | 2 (bases 1 to 2535) | | |
| REFERENCE | Park,C., Kawaguchi,T., Sumitani,J. and Arai,M. | | |
| AUTHORS | Direct Submission | | |
| TITLE | Submitted (15-JAN-2002) Motoo Arai, Osaka Prefecture University, Graduate School of Agriculture and Biological Sciences, Gakuen-cho 1-1, Sakai, Osaka 599-8531, Japan | | |
| JOURNAL | (E-mail:motoo@biochem.osaka-fu-u.ac.jp, URL:http://www.biochem.osaka-fu-u.ac.jp/AM/MAIN-J.html, Tel:81-72-254-9465(ex.2476), Fax:81-72-254-9921) | | |
| FEATURES | | | |
| source | Location/Qualifiers | | |
| | 1..2535 | | |
| | /organism="Streptomyces sp. M23" | | |
| | /strain="M23" | | |
| | /db_xref="taxon:184457" | | |
| | /clone_11b="pUC 118 Streptomyces sp. M23 genomic DNA" | | |
| | /country="Japan: Osaka, MIZUMA" | | |
| | 224..228 | | |
| | /evidence=not_experimental | | |
| | 241..1602 | | |
| | /gene="cbhi" | | |
| | /gene="cbhi" | | |
| | 241..330 | | |
| | /gene="cbhi" | | |
| | /evidence=experimental | | |
| | 241..1602 | | |
| | /gene="cbhi" | | |
| | /standard_name="cellobiohydrolase" | | |
| | /EC_number="3.2.1.91" | | |
| | /transl_start=1 | | |
| | /codon_start=1 | | |
| | /evidence=experimental | | |
| | /product="exoglucanase CBHI" | | |
| | /protein_id="BAB83928.1" | | |
| | /db_xref="GI:18250951" | | |
| | /translation="MSRSRTMLAALTLAAGSMTLALAAGPAGAAPAPARVDNPVYV GATVYVPEWSALASBPGDGRVADQPTAVALDIRIATIEGDKMGILREHLDEALQOR GSGELVQVLVYLDLGRDCAALASNGELDPEDLRKSEYIDPIADLISDKYGLRQ VHTWEPDPLPLVYNAGSDPTTEAGCTTKRANGVYEGGVSATSKLGAIPVYATIDA AHHGKMGDNLGSPVOEYFKVATSNKASVDVAGFVNTNANPSPTYEPPTVSDIVYN GQYRQSKWMDWMDNYVDEQSTQAQLRNEAFAAGNSDYGVIITSRMGWGSDRPSG GPGSVDAIYVDSRIDRRVHGVNMCNSGAGLGRPPAAPASGIDATWTIKLPESDGG NSAPVDDDEKGFDMQCDPSYOGNARNGVPSGLPPAPLPSGOWFSQFRELMOAYRP PLS" | | |
| | 16295 | | |
| terminator | /evidence=not_experimental | | |
| BASE COUNT | 386 a 1031 c 823 g 295 t | | |
| ORIGIN | | | |

[illegible]

| QY | 1099 | gtcaaacagtcgaagaattgacatttcgcaagcaccgagccgtgtgtgtcgaacagaacggt | 1158 |
|-----------------------|---|---|----------------------------|
| Db | 1270 | gttcgacggagacggccgcatgcacgaggccgctgccagtcggcgaactggttgcaaacagagccgc | 1329 |
| QY | 1159 | gtcgagccttcgagcagccgagcaggaagccagccagatcttcggaacgagcacttcgac | 1218 |
| Db | 1330 | gccccgcttcggc-----gacggccacaccccgccgcttcgggattccgattccgac | 1377 |
| QY | 1219 | gcgtatgtctgatacaagccgagctgtgagtcgacgagcacaacagcgtctgcagatccg | 1278 |
| Db | 1378 | gccctacaccttgatgaacggccggggagagtcgacagcagcaacagcggcccccgtcacaac | 1437 |
| QY | 1279 | acaactggaagaatctggaagcccaattgtgcgaccccgaaagtaagaagcttcgaagggtta | 1338 |
| Db | 1438 | gacgaaggcgaaggcttcgacccagatgtgcgaccccttcgtaccagagcgaaccccgcaac | 1497 |
| QY | 1339 | ctgacacaac-----gcgttacggaacatcccgatctcgagccgagctgtctccgag | 1389 |
| Db | 1498 | gggtgacaaacccctcggggttcgctgcccgacgacaccgctctccgggcaattggtttccgcc | 1557 |
| QY | 1300 | caagttgaccaagctgttcgcggaacacagccagcagctgtgcgaagctgcgaacagctgaagc | 1449 |
| Db | 1558 | cagttccggagactgattgacgaacccctaacccgctgctc-----ctgacccggtgaggggt | 1615 |
| QY | 1450 | ccgagcccttcgcgcgcgcgcgcgttcgcgcgttcgcgcgcgcgcgcgcgcgcgcgcgcgcgc | 1509 |
| Db | 1616 | ctttgaggtgattccggccggggcgggccggccggcggagccttcgcccctgactacccg | 1675 |
| QY | 1510 | agcagctgcacatcgccgctgc | 1569 |
| Db | 1676 | gccggcccccggggagacccgctgctccgcccggccgaccccgccggccggccgacgacaccc | 1735 |
| QY | 1570 | agtcggaagccgagtagctgc | 1629 |
| Db | 1736 | gccggccgggacacagccgctgctgctgctccggccggccggccggccggccggccggccgg | 1795 |
| QY | 1630 | agccgagatgc | 1689 |
| Db | 1796 | ggcacc--ggccgacgctgttcgaaagccggcgacacagacgccccttcgggcttctgcggc | 1853 |
| QY | 1690 | tcgacgagcccgagctgc | 1734 |
| Db | 1854 | tccccggcgacgactgcccacacacccggctgcccgttcgacgct | 1898 |
| RESULT | 5 | | |
| LOCUS | 169280 | 3404 bp | DNA linear PAT 04-FEB-1998 |
| DEFINITION | Sequence 1 from patent US 5677151. | | |
| ACCESSION | 169280 | | |
| VERSION | 169280.1 | GI:2831402 | |
| KEYWORDS | | | |
| SOURCE | Unknown. | | |
| ORGANISM | Unknown. | | |
| REFERENCE | 1 (bases 1 to 3404) | | |
| AUTHORS | Wilson,D.B., Walker,L.P. and Zhang,S. | | |
| TITLE | Thermotable cellulase from a thermomonospora gene | | |
| JOURNAL | Patent: US 5677151-A 1 14-OCT-1997; | | |
| FEATURES | Location/Qualifiers | | |
| source | 1..3404 | | |
| BASE COUNT | 619 a 1238 c 1005 g 542 t | | |
| ORIGIN | | | |
| Query Match | 8.8%; Score 324; DB 6; Length 3404; | | |
| Best Local Similarity | 56.6%; Pred. No. 1,2e-22; | | |
| Matches | 775; Conservative 0; Mismatches 553; Indels 60; Gaps 7; | | |
| QY | 200 | tcgtcaaccggtactggcgcaagaagtaagagcgaagcggcgaacacagacaaatgcca | 259 |

| | | | |
|---|------|--|------|
| D | 1010 | TCGACAAACCCCTTGGAGGGCCGCAAGCTGTACTGAACCCGGTCTGTGGCCAAAGCCG | 1065 |
| Q | 260 | ctctcgcagcgaaatgcgcgctcglttccaacatattcgacgcgcgtctglatgacgcga | 319 |
| D | 1070 | CCGGTCGAGCCGGGGGCTTCCGGCGGTGSCCAACAGTTCACCGGTGCTGGTGAACCGTA | 1129 |
| Q | 320 | tcgctgcgatctaaacggtctcaacggtgcgaaccggtcttgacgacatatcttgcagcgccgc | 379 |
| D | 1130 | TCGGGCGCATATGAGGGGCAAGACAGCCCGCACCGCTCTCATGGGTGTGCCGACCAAC | 1189 |
| Q | 380 | t-----ctccagcagcaggaagaaaccaacccttgaaagtaattgagatttcattcag | 430 |
| D | 1190 | TGGAGAGAGCGCGTCCGCCAGTCCGGTGTGGCAACCCCGTCGACATCCAGTGTCACTTACA | 1249 |
| Q | 431 | atctgcgcggagcgcgacttcgcgcgtctgcctctccaacgacgcgaattgccttcgaacgag | 490 |
| D | 1250 | ACCTGCCCCGGCGGACTGTGGCGCGCGGTGGCTTCAAAGGTGAGCGT-----GGTCCG | 1309 |
| Q | 491 | caggtttgcagactatgaaagcgcagtatacgatctcgatttcgagatctctgagaat- | 549 |
| D | 1304 | ATGAACTCGACCGCTACAAAGAGAGTACATCCAGCCGATCCGCCATCATGTGGGACT | 1363 |
| Q | 550 | --cgaagtatccagccttcgcatgtgacgtatgcatacttgagccgcgacttcgcgaacg | 607 |
| D | 1364 | TCGCAGACTTACGAGAACCTCGGATGCTGCCCATCTCAAGATCGACTCTCCGCCCCAAC | 1423 |
| Q | 608 | cggtaccacatgatgattcaacagcgtgtgtcaacgacggtgcgtg-----653 | |
| D | 1424 | TCGTTCACCAAGTGTGGGGGAGACGGCGGCACCGACTCTGCGCTTCATGAAAGCAAGAC | 1483 |
| Q | 654 | -----ttagcgaacagatcgtgagtgacgcgtctcagaatttcagcgcattccgaacgtgt | 709 |
| D | 1484 | CGCGCTACGTCACAGGTGTGGCTGACGCCCTCCGCAAGCTGGGGCGAGTCCGAAACGTCT | 1543 |
| Q | 710 | acattctatgtagaacgcgcgcccttcgcgtcgtgtgtgtgtgcaccaatattgcagcgaat | 769 |
| D | 1544 | ACAACCTATATGACGGCCGCCACACGCGCTGAGTCCGGTGGGATCTCAACTTCCTGGCCCT | 1603 |
| Q | 770 | acgtacgaagaattcagagaagtcctcaacg---cgagatctggggttcaacgcgcatacgag | 826 |
| D | 1604 | CGGTGGACATTTTACGAGGGCCGCCAACGCCCTTCGGCTTCACCGTGGACTACGTGCACG | 1665 |
| Q | 827 | gcttcgtcaaccaacacgcgcgaattcaacgcgcgtttgaagagacgcgttcatgacgcgcaccc | 886 |
| D | 1684 | GCTTCATCTCCAACACAGGCCCACTACCTCGGCACTGTGGAGCGGTACTCGAGACGTCAAG | 1722 |
| Q | 887 | agcaggtctgcgcgcgaacgcgcggttgagtgctgcgcgaatttctaacagtgtgaattcctgacatg | 946 |
| D | 1724 | GCACCGTTTAAAGGGCACCTCATCCCGCAGTCCAAAGTGGGTGATCGTAACAGTACGTGTC | 1783 |
| Q | 947 | acgaagccgactttagcgcgttgacttgaacttcgcgcgcggtctgcgcgtcttccaaga | 1006 |
| D | 1784 | ACGAGCTTCTCTTGTTCACAGGACTGTGCTAGGCCCTTGATTCGCNAAGGCTTCCGGTCCG | 1843 |
| Q | 1007 | gcaticgcgatctcatcgacaaccttaacgaacggtttgggtgtgtccgaacgaacgaacag | 1066 |
| D | 1844 | ACATCGGATCTCATGACACACCTCCCGCAAGGCGTGGGGTGGCCGAAACCTCCGACCG | 1903 |
| Q | 1067 | gcccgagcagcgcgacgcgagtcaacaaccttgtaacaaacagtcgaagaattgacattgcgc | 1126 |
| D | 1904 | GACCGAGCTCTCTCCACCGACCTCAACACCTTACACAGAGACCCGTATTCGACCCGCTTA | 1965 |
| Q | 1127 | agcagcgcgcgtcgttggttgaaacagaaagcgtgcgcgccttcgcgcgaacgcgcgaagcga | 1186 |
| D | 1964 | TCACACCCCGTAACTGTGTGCAACCAAGCCGCGTCCGGGCTCCGGCGAGC----- | 2011 |
| Q | 1187 | gcccgagcgaattctccgaacgcgcgaaccttcgaacgcgtatgtctgatacgaacgcgcgcggtgt | 1246 |
| D | 2012 | GGCCCAACGCTAAACCCGGCTCCCGGTGTTCACGCCCTACGTCTGTGTGGAAGCCCCCGGGTG | 2071 |
| Q | 1247 | agtcgcagcagacaagcgtctgcgaagatctcgacaaacttgcgaagaagtgcgaaccatgtct | 1306 |
| D | 2072 | AGTCCGAGCGGCCACGAGAGAGTCCGCAACACAGAGGGCAAGGGCTTCGACCCGCACTGT | 2131 |

| | | | |
|--------------|--|--|----------------------------|
| Oy | 1307 | gagaccagcagctaacgcagcagctgcgtacgggagctacgtaccac-----ggctaacgga | 1357 |
| Db | 2132 | GGGAGCCCAACCTCTAAAGCGGAGCAACGCCCCCAACGCGCAACACCCCTCGGCTGGCGTGC | 2191 |
| Oy | 1358 | actcccgcatcgccgcgcagcttgcttccggcgcaagcttgacacagcttgctcgcaacgac | 1417 |
| Db | 2192 | ACGCCCCATCTCCGGCCACATGCTCTCTGTGCCAATTCCGGAGAGTGTGTCGCAACGCTT | 2251 |
| Oy | 1418 | ggccagcggtgtgcgcagcgtcgacacagctcgagccgcgcgccttcgcgcgcgcagcttcg | 1477 |
| Db | 2252 | ACCCCTCTGTAAAGCGGAGTGAAGCAACGAGCTGACAGCTCAACGAGGAACGTATCAG | 2311 |
| Oy | 1478 | cttcgcgcagctcgagccgcgcagctcgagccgcgcgcgcgcgcgcgcgcgcgcgcgcgc | 1527 |
| Db | 2312 | CACCTCTTACGCGGAGACGCGCCCTGTCACCTCCCGTGGGCGGCGCGC | 2361 |
| RESULT | 6 | | |
| LOCUS | TFU18978 | 3503 bp | DNA linear BCT 13-NOV-2001 |
| DEFINITION | Thermomonospora fusca beta-1,4-exocellulase precursor gene, | | |
| ACCESSION | U18978 | | |
| VERSION | U18978.1 | GI:664822 | |
| KEYWORDS | Thermobifida fusca. | | |
| SOURCE | Thermobifida fusca. | | |
| ORGANISM | Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptosporangineae; Thermomonosporaceae; Thermobifida. | | |
| REFERENCE | 1 (bases 1 to 3503) | | |
| AUTHORS | Zhang, S., lao, G. and Wilson, D. B. | | |
| TITLE | Characterization of a Thermomonospora fusca exocellulase | | |
| JOURNAL | Biochemistry 34 (10), 3386-3395 (1995) | | |
| MEDLINE | 95186496 | | |
| PUBMED | 7880834 | | |
| REFERENCE | 2 (bases 1 to 3503) | | |
| AUTHORS | Zhang, S. | | |
| TITLE | Direct Submission | | |
| JOURNAL | Submitted (21-DEC-1994) Sheng Zhang, Biochemistry, Molecular & Cell Biology, Cornell University, 460 Biotechnology Building, Ithaca, NY 14853, USA | | |
| FEATURES | Location/Qualifiers | | |
| source | 1..3503 | | |
| | /organism="Thermobifida fusca" | | |
| | /strain="YX" | | |
| | /db_xref="taxon:2021" | | |
| | /clone="ps24; ps26" | | |
| | 249..261 | | |
| protein_bind | /gene="E3" | | |
| | /bound_moiety="regulatory protein" | | |
| gene | 249..3506 | | |
| | /gene="E3" | | |
| 5' UTR | 262..574 | | |
| | /gene="E3" | | |
| stem_loop | 349..383 | | |
| | /gene="E3" | | |
| protein_bind | 350..363 | | |
| | /gene="E3" | | |
| | /bound_moiety="regulatory protein" | | |
| stem_loop | 467..504 | | |
| | /gene="E3" | | |
| RBS | 560..564 | | |
| | /gene="E3" | | |
| CDS | 575..2365 | | |
| | /gene="E3" | | |
| | /EC_number="3.2.1.91" | | |
| | /note="Cel C precursor" | | |
| | /codon_start=1 | | |
| | /transl_table=11 | | |
| | /product="beta-1,4-exocellulase precursor" | | |
| | /protein_id="AA62211.1" | | |

[illegible]

/gene="SC5C7.02", uvra-1-like protein, len: 796 aa; similar to Streptomyces peucedis NR_054827 (EMBL:L76359) daunorubicin resistance protein drbc (764 aa), fasta scores; opt: 2051 z-score: 2702.2 E(): 0, 51.1% identity in 775 aa overlap, and to many uvra proteins e.g. UVRA_NEI00 excinuclease ABC subunit A (950 aa), fasta scores; opt: 412 z-score: 1697.0 E(): 0, 34.1% identity in 933 aa overlap. Contains 2x P500017 ATP/GTP-binding site motif A (P-loop) and 2x Pfam match to entry PF00005 ABC_tran, ABC transporters, scores 42.40, E-value 1.7e-10 and 71.20, E-value 2.2e-17"
/codon_start=-1
/transl_table=11
/product="uvra-1-like protein"
/protein_id="CAA20614.1"
/db_xref="GI:3559992"
/translation="MSSAKRPGTGGSHVADSHD.LRVHGAARENNLKDVSVDPKRR LTVFPGVSSSKSLVFTNTIAESQRLLNETYSAPVGCFPTLRLAPEVDIDGITNLN IDDOORGMADPSTVIGATNDVANMLRI.FSLRGEPRIGPPSAYSFMTASVASCALITV ENGNKRVAATEERSTGTCCTHC.EGRGVSDYLDTQLXDSDSNLAGAFTIPGMSDSOS MIVQYVAGSGFEDPKPIREYTEKEELRFPLGEPEPVKNVNLTIEBELLIOKSEFL SKDFKAMOPHKAIFERAVALTFTECEGTCTEGBARSX.IKKTSIADCAAME.RDLA EMVRILTFSVPALILTLARDLSPVEGLG.TSIDRAPGCTSGEENQRYMYTHLSS SITDTTYTDEPTVLPHHDIORAMDILLRLDKQNTLYLVENHP.EALAIADHVDDVLD PAGTAGTGVCETGETVELRADVTIGHHLDRDLAVLKSVSRPA.GALEIRPARHTNI.Q LPDINCVLGLCLVTGVAGSGSKSLIHCSVPAGADVVSDDSPIKGSRSNPATYTGLI GDIPIRKARPAKANGVAF.PLFANSEBACPTCGAGITYTLDLMAGVATPCDCGKRKR OPAVLEVERGGRDISEVLA.MSYDOEEFGEGEARTPAAKH.IIOLRISVGLGTLICQ PLTTISGGEROBLKIATHTMGEXGGVYV.DEPTGTHLADVOLLGLDRLVDARKSVY VIEHHQAVAHADM.IIIDGPGAGHDGGRVTEGIPADLVADRSTLTGEHLAAYGA " 411..1700
/gene="SC5C7.02"
/note="Pfam match to entry PF00005 ABC_tran, ABC transporters, score 42.40, E-value 1.7e-10"
492..515
/gene="SC5C7.02"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
1893..2627
/gene="SC5C7.02"
/note="Pfam match to entry PF00005 ABC_tran, ABC transporters, score 71.20, E-value 2.2e-17"
1914..1937
/gene="SC5C7.02"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
2803..3285
/gene="SC5C7.03"
2803..3285
/gene="SC5C7.03"
/note="SC5C7.03, probable transcriptional regulator, len: 160 aa; similar in N-terminus to many e.g. RAMA_ENRCL transcriptional activator rama (113 aa), fasta scores; opt: 175 z-score: 255.3 E(): 5.7e-07, 32.3% identity in 96 aa overlap. C-terminal 50 aa are unique. Contains Pfam match to entry PF00165 HTH_2, Bacterial regulatory helix-turn-helix proteins, arac family, score 71.40, E-value 1.9e-17"
/codon_start=-1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="CAA20615.1"
/db_xref="GI:3559993"
/db_xref="SPTREMBL:086700"
/translation="MTLEEDLARARADDMDREYAEPDYDPALARVALMSGAFHSR FFAANGEPYSTLTMRTRIERARALLRCDMSVTECFVAGTSLGSSFSTRTELVEGES PSAYERHRDADGALPACVAKYTRPVANGEGRGTPAGSSESVRN.EAAPAPPVPA " 2872..3132
/gene="SC5C7.03"
/note="Pfam match to entry PF00165 HTH_2, Bacterial regulatory helix-turn-helix proteins, arac family, score 71.40, E-value 1.9e-17"

```

gene      3291.. 3707
          /gene="SC5C7.04"
CDS       3291.. 3707
          /gene="SC5C7.04"
          /note="SC5C7.04, possible lyase, similar to e.g.
          IGUL.HAEM lactoylglytathione lyase (EC 4.4.1.5) (135 aa)
          fasta scores: opt:124 z-score: 215.2 E(): 9.7e-05, 29.0%
          identity in 131 aa overlap"
          /codon_start=1
          /product="putative lyase"
          /protein_id="CAA20616.1"
          /db_xref="GI:3559994"
          /db_xref="SPTREMBL:O86701"
RBS       /translation="MDISLSQCFIAVDHDDBALAFRDVLCLEVRNDYGYGMNRVTIT
GSPAQPDVIDILEPPLADNPASADKAMADLLKGLRGIIILTDVDAPFEHVRRAA
GAELVQEPDIPYCVRDCAFPRDPAGNMRHLRPKG"
          3978.. 3981
          /note="possible RBS upstream of SC5C7.05"
          3994.. 4794
          /gene="SC5C7.05"
          3994.. 4794
          /gene="SC5C7.05"
          3994.. 4794
          /gene="SC5C7.05"
          /note="SC5C7.05, probable RNA polymerase sigma factor,
          len: 266 aa; similar to S. coelicolor sigma factors of the
          sigB subfamily e.g. TR:0523j3 (EMB1:AR036131) putative RNA
          polymerase secondary sigma factor sigC (263 aa), fasta
          scores: opt: 1258 z-score: 1540.5 E(): 0, 72.9% identity
          in 251 aa overlap, and to RPof-STRAU RNA polymerase
          sigma-F factor (297 aa), fasta scores: opt: 713 z-score:
          858.3 E(): 0, 42.6% identity in 312 aa overlap. Possible
          alternative start 216 bp upstream, not supported by
          Frameplot. Contains helix-turn-helix motif at aa 228-249
          (Score 2104, +6.35 SD) and Pfam match to entry PF00140
          sigma70. Sigma-70 factors, score 210.40, E-value 2.8e-59"
          /codon_start=1
          /transl_table=11
          /product="putative RNA polymerase sigma factor"
          /protein_id="CAA20617.1"
          /db_xref="GI:3559995"

Query Match      8.7%; Score 320; DB 1; Length 41906;
Best Local Similarity 55.5%; Pred. No. 1.3e-22;
Matches 773; Conservative 0; Mismatches 565; Indels 54; Gaps .6.
```


[illegible]

| Accession | Version | GI | Accession | Version | GI |
|---------------------------------------|---------------------------------------|------------|---------------------------------------|---------------------------------------|------------|
| 169283 | 169283.1 | GI:2831405 | 169283 | 169283.1 | GI:2831405 |
| Keywords | Keywords | | Keywords | Keywords | |
| Source | Source | | Source | Source | |
| Organism | Organism | | Organism | Organism | |
| Reference | Reference | | Reference | Reference | |
| Authors | Authors | | Authors | Authors | |
| Title | Title | | Title | Title | |
| Journal | Journal | | Journal | Journal | |
| Features | Features | | Features | Features | |
| Source | Source | | Source | Source | |
| Base Count | Base Count | | Base Count | Base Count | |
| Origin | Origin | | Origin | Origin | |
| Query Match | Query Match | | Query Match | Query Match | |
| Best Local Similarity | Best Local Similarity | | Best Local Similarity | Best Local Similarity | |
| Matches 732; Conservative | Matches 732; Conservative | | Matches 732; Conservative | Matches 732; Conservative | |
| 8.5%; Score 314.6; DB 6; Length 1269; | 8.5%; Score 314.6; DB 6; Length 1269; | | 8.5%; Score 314.6; DB 6; Length 1269; | 8.5%; Score 314.6; DB 6; Length 1269; | |
| 57.6%; Pred. No. 1.2e-21; | 57.6%; Pred. No. 1.2e-21; | | 57.6%; Pred. No. 1.2e-21; | 57.6%; Pred. No. 1.2e-21; | |
| 0; Mismatches 479; Indels 60; Gaps | 0; Mismatches 479; Indels 60; Gaps | | 0; Mismatches 479; Indels 60; Gaps | 0; Mismatches 479; Indels 60; Gaps | |
| 7/ | 7/ | | 7/ | 7/ | |
| 200 | 200 | | 200 | 200 | |
| 17 | 17 | | 17 | 17 | |
| 260 | 260 | | 260 | 260 | |
| 77 | 77 | | 77 | 77 | |
| 320 | 320 | | 320 | 320 | |
| 137 | 137 | | 137 | 137 | |
| 380 | 380 | | 380 | 380 | |
| 197 | 197 | | 197 | 197 | |
| 431 | 431 | | 431 | 431 | |
| 257 | 257 | | 257 | 257 | |
| 491 | 491 | | 491 | 491 | |
| 311 | 311 | | 311 | 311 | |
| 548 | 548 | | 548 | 548 | |
| 371 | 371 | | 371 | 371 | |
| 608 | 608 | | 608 | 608 | |
| 431 | 431 | | 431 | 431 | |
| 655 | 655 | | 655 | 655 | |
| 491 | 491 | | 491 | 491 | |
| 710 | 710 | | 710 | 710 | |
| 551 | 551 | | 551 | 551 | |
| 770 | 770 | | 770 | 770 | |
| 611 | 611 | | 611 | 611 | |
| 837 | 837 | | 837 | 837 | |
| 671 | 671 | | 671 | 671 | |
| 887 | 887 | | 887 | 887 | |
| 731 | 731 | | 731 | 731 | |

Db 824 GCACCTTTAAAGCGCAGCTCATCCGCCAGTCCAGAGTGCTGACTGAGACCACTAGCTCG 883
Qy 947 acgaagcgcagctacgagcggttgaacttgtaactgctgcgcgcgcgcgttcccaagca 1006
Db 884 ACAGAGCTCCTTCCTGTCAGAGAGCTCGCTACAGCCCTCGATGCGCAAGCGCTTCGGCTCCG 943
Qy 1007 gcatcgagctatcgtacacacctaagcaagcttgggggtgtgtccgaacgaacaaag 1066
Db 944 ACATCGGATGATGCTATGAGACACTCCCGCAACGGCTGGGGTGGCCGCAACCTGCCAGCG 1003
Qy 1067 gcccgagcagcgcagcagcagatgttaacacctgttcaaccgctgaagatgtacctcgcc 1126
Db 1004 GACCGAGCTCTCTCCAGCAGACTCACTACCTGACGAGACCGCATGATGACCGCGTA 1063
Qy 1127 agcagcgcgcgcctgtgtgtaacacagaagctgagcgcgcgcgcgcgcgcgcgcgcgcgc 1186
Db 1064 TCCACCCCGGTATCTGTGTAACACCAAGCCCGCTGCGGCTCGGCGAGCGGCC----- 1116
Qy 1187 gcccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1246
Db 1117 -----ACGGCAACCCGCGCTCCGGTGTGACCGCTACGCTGCTGGTGAACCCCGGGTGC 1171
Qy 1247 agtcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1306
Db 1172 AGTCCGACGGCGCGCAGAGAGATCCCGAACGACGAGGCGCAAGGCGCTTGACCGCATGT 1231
Qy 1307 gccagcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1357
Db 1232 GCGACCCGACACTACGAGGGAACGCGCCGCAACGCGCAACACCCCTCGGGTGGCTGCGCA 1291
Qy 1358 actcccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1417
Db 1292 ACGCCCCCACTCCGCGCAGCTGCTCTGCTCCAGCTTCCGAGCTGCTGCGCAACGCGCT 1351
Qy 1418 ggcagcagcagc 1428
Db 1352 ACCCGCTCTCG 1362

RESULT 10
AL646079
LOCUS 213050 bp DNA linear BCT 07-DEC-2001
DEFINITION Ralstonia solanacearum GMI1000 megaplasmid, complete sequence;
ACCESSION AL646079 AL646053
VERSION
KEYWORDS AL646079.1 GI:17430956
SOURCE Ralstonia solanacearum.
ORGANISM Ralstonia solanacearum
Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
Ralstonia.
1 (bases 1 to 213050)
Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S.,
Arlet,M., Billault,A., Brotier,P., Camus,J.C., Catolico,L.,
Chandler,M., Choisme,N., Claudel-Renard,C., Cunnac,S., Demange,N.,
Gaspin,C., Lavie,M., Moisan,A., Robert,C., Saurin,W., Schlex,T.,
Sigler,P., Thebaud,P., Whalen,M., Wincker,P., Levy,M.,
Weissenbach,J. and Boucher,C.A.
Genome sequence of the plant pathogen Ralstonia solanacearum
Unpublished
2 (bases 1 to 213050)
Boucher,C.A.
Direct Submission
Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,
BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
Dausset-CRPH, 27 rue Juliette Dodu, 75010 Paris, France, IMGM CNRS
118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,
Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,
F31326 Castanet-Tolosan Cedex, Laboratoire de Genetique Cellulaire
INRA, BP27, F31326 Castanet-Tolosan Cedex

COMMENT Christian.Boucher@toulouse.inra.fr
FEATURES http://sequence.toulouse.inra.fr/R.solanacearum.html.
location/Qualifiers
1. 213050
/organism="Ralstonia solanacearum"
/plasmid="megaplasmid"
/strain="GMI1000"
/db_xref="taxon:305"
100. .1632
/gene="RSP0487"
/note="RS00365"
100. .1632
/gene="RSP0487"
/function="macromolecule metabolism; macromolecule
degradation; degradation of proteins, peptides,
glycopeptides"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="HYPOTHETICAL TRANSMEMBRANE PROTEIN"
/protein_id="CAD17638.1"
/db_xref="GI:17430957"
/transl_table="MDPSCRGAVIRAVRGRLLVLYVGLAVLCIAFNTRHDSR
QIDVPPAPLAVGAAARIGCAITARTISTTDQONADQROUHAMLAARYPAH
AAQRPEPVDLSLYTWKSDPSLKLILMAHODVVPVARGTEGDMTEPEFACVADG
MWGRAMDKGNLIOMEAEALIASGEPRIITHFAGADEVGEERAAIALIL
KSRGEOLAEVIDEGLITEGVVPLAIPALIVTEKGLSVALKIAPGHSMPA
PGESAIAMMSAALKHIDDOOLPAGITGVREMEETLAPEMGRNVALSLMLFEPLY
RKOIASPSTNAVLOTFTALFTVQAGNKQNVLPRAEATVYNEPRLPGDSASVIAVE
QAVRSAYPKGHEFLALPQVSEAPYSPPOSASYQILGRVREPFPGTVAPAPLMGA
TDSRHMIGISDHFERSVPAKPEDLAREHGTERTSEANVELIRFYHRLIHQADA
GA"
complement(1751. .3148)
/gene="czcs"
/note="RSP0488; RS00366"
complement(1751. .3148)
/gene="czcs"
/fc_number="2.7.3."
/function="cell processes; protection responses"
/note="Product confidence : probable
Gene name confidence : probable
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="PROBABLE TRANSMEMBRANE SENSORY TRANSDUCTION
HISTIDINE KINASE FOR COBALT ZINC CADMIUM RESISTANCE
TRANSCRIPTION REGULATOR PROTEIN"
/protein_id="CAD17639.1"
/db_xref="GI:17430958"
/transl_table="MMRPASLSAARLVAFVLYAGVLMGAGTYLYALNAOLRARD
AMLGKIVOLRHCAKDTQSIDIVPARRTEVSGHEGLITLVIPDGTLEAPNE
GEVPAAPVPPVPLASLSALHTWMSRTGASRAIAOARLGNHRTVTVLVADKDS
RAALMSYRHHIFLAVIAGVAVALSLMVRSTGLARIGQOASVPSQDTRLT
LEGAPRELSDILALNMLTRLQDGFARLSQFSAADLHDFRPLANLIGQETVLAHP
RPAGEYRAVESLSLEEARLSMIEDMLFLARDBARVALSVAPDLARHETLOLVDF
EAVAADRDTITOCGCGRYRADITLIRAINLIDNALRHTPGCATIDVAVOVLPSST
AIVARNPGEITPAASLPPLFDRYRQDPPARANSERAGLALIVOTIMDHGGTAQAS
SPGSGTLEFRLLPFA"
complement(3145. .3819)
/gene="czcr"
/note="RSP0489; RS00367"
complement(3145. .3819)
/gene="czcr"
/function="cell processes; protection responses"
/note="Product confidence : probable

gene
CDS

Gene name confidence : probable
Predicted by Codon_usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="PROBABLE RESPONSE REGULATOR FOR COBALT ZINC
CADMIUM RESISTANCE TRANSCRIPTION REGULATOR PROTEIN"
/protein_id="CAD1640.1"
/db_xref="GI:17430959"
/translation="MRLIVDEPRKAGDYLLKLTGSEGFVADLARSQPDGLYQMEND
VDLYLVDMVGOMGVITRELRRRTPTPVLEFARDELSPRLKGADDDYVKRPF
APAEIYAVALHIIILRGPMRESEVELEIDITDTIKRRVTGRGRIDITFAEYALHTLL
ARRGEVYSRLISSQWMDVNFDSNTNVDAVRIIRLRKAVDDPEHKLIHLKGMGTIV
LDYVAR"

4089..4394
/gene="RSP0490"
/note="RS00368"
4089..4394
/gene="RSP0490"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : putative
Gene name confidence : hypothetical
Predicted by Codon_usage
Predicted by Framed"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="PUTATIVE SIGNAL PEPTIDE PROTEIN"
/protein_id="CAD17641.1"
/db_xref="GI:17430960"
/translation="MSVRNCLPLPAAAFELACGAATASVSQALSPIQVHGARDP
YTDRKGRDRTYTGGAQEATIQASVGRDPYPDGMSRSSAPRDHPDYGQAQA"

4400..5725
/gene="czcc"
/note="RSP0491: RS00369"
4400..5725
/gene="czcc"
/function="cell processes; protection responses"
/note="Product confidence : probable
Gene name confidence : probable
Predicted by Homology
Predicted by Codon_usage
Predicted by Framed"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="PROBABLE COBALT-ZINC-CADMIUM OUTER MEMBRANE
RESISTANCE PROTEIN"
/protein_id="CAD17642.1"
/db_xref="GI:17430961"
/translation="MTHTPCIRAAVAVLSICALTVSRGLAAPAGEMAASPATAPAOS
APASPGAAHHMRITLQALIDRAIALTHPALAAGYAVASBGAVDAQSLRNPREAFLTT
EDVNRDRTVTRQLNTIPLEGKRRAARTRAETLRDAGDPAVAIAELRDVYTRAFF
ELISAOERLRLAQSDALAAQAEAAAVRVQAGVSPVEQKAAVARAGAIELRDQAQ
ADLTAIEQALSAFWGGAVAPEADGSATTLPVAPPEAHRYLAATAEPALASTPPAQO
LRAELEIHROALTVNGVERSKRPVPTVTSVCAKRDGSANNAVVGIAPVPIFRONG
SILEORLADAEDYRALRIEQAALISODVARLEAAROAVGALRODVTLGCAGAYTA
ARIGEAACKRFPLVDLAQRILTFOARAQYTLAISRHAQAAICDIRITGR"

5752..6966
/gene="czcb"
/note="RSP0492: RS00370"
5752..6966
/gene="czcb"
/function="cell processes; protection responses"
/note="Product confidence : probable
Gene name confidence : probable
Predicted by Codon_usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/transl_table=11

gene
CDS

Gene name confidence : probable
Predicted by Codon_usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="PROBABLE RESPONSE REGULATOR FOR COBALT ZINC
CADMIUM RESISTANCE TRANSCRIPTION REGULATOR PROTEIN"
/protein_id="CAD1640.1"
/db_xref="GI:17430959"
/translation="MRLIVDEPRKAGDYLLKLTGSEGFVADLARSQPDGLYQMEND
VDLYLVDMVGOMGVITRELRRRTPTPVLEFARDELSPRLKGADDDYVKRPF
APAEIYAVALHIIILRGPMRESEVELEIDITDTIKRRVTGRGRIDITFAEYALHTLL
ARRGEVYSRLISSQWMDVNFDSNTNVDAVRIIRLRKAVDDPEHKLIHLKGMGTIV
LDYVAR"

4089..4394
/gene="RSP0490"
/note="RS00368"
4089..4394
/gene="RSP0490"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : putative
Gene name confidence : hypothetical
Predicted by Codon_usage
Predicted by Framed"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="PUTATIVE SIGNAL PEPTIDE PROTEIN"
/protein_id="CAD17641.1"
/db_xref="GI:17430960"
/translation="MSVRNCLPLPAAAFELACGAATASVSQALSPIQVHGARDP
YTDRKGRDRTYTGGAQEATIQASVGRDPYPDGMSRSSAPRDHPDYGQAQA"

4400..5725
/gene="czcc"
/note="RSP0491: RS00369"
4400..5725
/gene="czcc"
/function="cell processes; protection responses"
/note="Product confidence : probable
Gene name confidence : probable
Predicted by Homology
Predicted by Codon_usage
Predicted by Framed"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="PROBABLE COBALT-ZINC-CADMIUM OUTER MEMBRANE
RESISTANCE PROTEIN"
/protein_id="CAD17642.1"
/db_xref="GI:17430961"
/translation="MTHTPCIRAAVAVLSICALTVSRGLAAPAGEMAASPATAPAOS
APASPGAAHHMRITLQALIDRAIALTHPALAAGYAVASBGAVDAQSLRNPREAFLTT
EDVNRDRTVTRQLNTIPLEGKRRAARTRAETLRDAGDPAVAIAELRDVYTRAFF
ELISAOERLRLAQSDALAAQAEAAAVRVQAGVSPVEQKAAVARAGAIELRDQAQ
ADLTAIEQALSAFWGGAVAPEADGSATTLPVAPPEAHRYLAATAEPALASTPPAQO
LRAELEIHROALTVNGVERSKRPVPTVTSVCAKRDGSANNAVVGIAPVPIFRONG
SILEORLADAEDYRALRIEQAALISODVARLEAAROAVGALRODVTLGCAGAYTA
ARIGEAACKRFPLVDLAQRILTFOARAQYTLAISRHAQAAICDIRITGR"

5752..6966
/gene="czcb"
/note="RSP0492: RS00370"
5752..6966
/gene="czcb"
/function="cell processes; protection responses"
/note="Product confidence : probable
Gene name confidence : probable
Predicted by Codon_usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/transl_table=11

| Query Match | Best Local Similarity | Score | DB 1: | Length | DB 2: | Length | DB 3: | Length | DB 4: | Length | DB 5: | Length | DB 6: | Length | DB 7: | Length | DB 8: | Length | DB 9: | Length | DB 10: | Length | DB 11: | Length | DB 12: | Length | DB 13: | Length | DB 14: | Length | DB 15: | Length | DB 16: | Length | DB 17: | Length | DB 18: | Length | DB 19: | Length | DB 20: | Length | DB 21: | Length | DB 22: | Length | DB 23: | Length | DB 24: | Length | DB 25: | Length | DB 26: | Length | DB 27: | Length | DB 28: | Length | DB 29: | Length | DB 30: | Length | DB 31: | Length | DB 32: | Length | DB 33: | Length | DB 34: | Length | DB 35: | Length | DB 36: | Length | DB 37: | Length | DB 38: | Length | DB 39: | Length | DB 40: | Length | DB 41: | Length | DB 42: | Length | DB 43: | Length | DB 44: | Length | DB 45: | Length | DB 46: | Length | DB 47: | Length | DB 48: | Length | DB 49: | Length | DB 50: | Length | DB 51: | Length | DB 52: | Length | DB 53: | Length | DB 54: | Length | DB 55: | Length | DB 56: | Length | DB 57: | Length | DB 58: | Length | DB 59: | Length | DB 60: | Length | DB 61: | Length | DB 62: | Length | DB 63: | Length | DB 64: | Length | DB 65: | Length | DB 66: | Length | DB 67: | Length | DB 68: | Length | DB 69: | Length | DB 70: | Length | DB 71: | Length | DB 72: | Length | DB 73: | Length | DB 74: | Length | DB 75: | Length | DB 76: | Length | DB 77: | Length | DB 78: | Length | DB 79: | Length | DB 80: | Length | DB 81: | Length | DB 82: | Length | DB 83: | Length | DB 84: | Length | DB 85: | Length | DB 86: | Length | DB 87: | Length | DB 88: | Length | DB 89: | Length | DB 90: | Length | DB 91: | Length | DB 92: | Length | DB 93: | Length | DB 94: | Length | DB 95: | Length | DB 96: | Length | DB 97: | Length | DB 98: | Length | DB 99: | Length | DB 100: | Length | DB 101: | Length | DB 102: | Length | DB 103: | Length | DB 104: | Length | DB 105: | Length | DB 106: | Length | DB 107: | Length | DB 108: | Length | DB 109: | Length | DB 110: | Length | DB 111: | Length | DB 112: | Length | DB 113: | Length | DB 114: | Length | DB 115: | Length | DB 116: | Length | DB 117: | Length | DB 118: | Length | DB 119: | Length | DB 120: | Length | DB 121: | Length | DB 122: | Length | DB 123: | Length | DB 124: | Length | DB 125: | Length | DB 126: | Length | DB 127: | Length | DB 128: | Length | DB 129: | Length | DB 130: | Length | DB 131: | Length | DB 132: | Length | DB 133: | Length | DB 134: | Length | DB 135: | Length | DB 136: | Length | DB 137: | Length | DB 138: | Length | DB 139: | Length | DB 140: | Length | DB 141: | Length | DB 142: | Length | DB 143: | Length | DB 144: | Length | DB 145: | Length | DB 146: | Length | DB 147: | Length | DB 148: | Length | DB 149: | Length | DB 150: | Length | DB 151: | Length | DB 152: | Length | DB 153: | Length | DB 154: | Length | DB 155: | Length | DB 156: | Length | DB 157: | Length | DB 158: | Length | DB 159: | Length | DB 160: | Length | DB 161: | Length | DB 162: | Length | DB 163: | Length | DB 164: | Length | DB 165: | Length | DB 166: | Length | DB 167: | Length | DB 168: | Length | DB 169: | Length | DB 170: | Length | DB 171: | Length | DB 172: | Length | DB 173: | Length | DB 174: | Length | DB 175: | Length | DB 176: | Length | DB 177: | Length | DB 178: | Length | DB 179: | Length | DB 180: | Length | DB 181: | Length | DB 182: | Length | DB 183: | Length | DB 184: | Length | DB 185: | Length | DB 186: | Length | DB 187: | Length | DB 188: | Length | DB 189: | Length | DB 190: | Length | DB 191: | Length | DB 192: | Length | DB 193: | Length | DB 194: | Length | DB 195: | Length | DB 196: | Length | DB 197: | Length | DB 198: | Length | DB 199: | Length | DB 200: | Length | DB 201: | Length | DB 202: | Length | DB 203: | Length | DB 204: | Length | DB 205: | Length | DB 206: | Length | DB 207: | Length | DB 208: | Length | DB 209: | Length | DB 210: | Length | DB 211: | Length | DB 212: | Length | DB 213: | Length | DB 214: | Length | DB 215: | Length | DB 216: | Length | DB 217: | Length | DB 218: | Length | DB 219: | Length | DB 220: | Length | DB 221: | Length | DB 222: | Length | DB 223: | Length | DB 224: | Length | DB 225: | Length | DB 226: | Length | DB 227: | Length | DB 228: | Length | DB 229: | Length | DB 230: | Length | DB 231: | Length | DB 232: |
|-------------|-----------------------|-------|-------|--------|-------|--------|-------|--------|-------|--------|-------|--------|-------|--------|-------|--------|-------|--------|-------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|
|-------------|-----------------------|-------|-------|--------|-------|--------|-------|--------|-------|--------|-------|--------|-------|--------|-------|--------|-------|--------|-------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|

| | | | |
|---|------|--|--|
| QY | 739 | tgagcttggtgagcccaataatgccaagcgatcagctacaggaatgcacaaagttcctcaac | 798 |
| Db | 883 | TGGCTCGGCTGGGACTCGACCGGGCTCCGGCGGAGAGATCTTCGCCGACCAAG | 942 |
| QY | 799 | gcgcagcatcggtgatacgcg---gcatacgagcgcttcgtcaccaacagcggaattacacg | 855 |
| Db | 943 | ACGACCAACGGGGGCTTCGCTCGATCACCGGCTTCGTCGAGCGATCGCACACACG | 1002 |
| QY | 856 | ccgctgaagagacc | 869 |
| Db | 1003 | CCGCTTGAGAGACC | 1016 |
| RESULT | 12 | | |
| AE003960/c | | 11541 bp | DNA linear BCT 15-JUN-2001 |
| LOCUS | | | |
| DEFINITION | | | Xylella fastidiosa 9a5c, section 106 of 229 of the complete genome. |
| VERSION | | | AE003960 AE003849 |
| KEYWORDS | | | AE003960.1 GI:9106242 |
| SOURCE | | | |
| ORGANISM | | | Xylella fastidiosa 9a5c. Xylella fastidiosa 9a5c Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group; Xylella. |
| REFERENCE | | | |
| ATTNORS | | | |
| 1 (bases 1 to 11541) | | | |
| Simpson,A.J., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M., | | | |
| Alvarenga,R., Alves,L.M., Araya,J.E., Bala,G.S., Baptista,C.S., | | | |
| Bartos,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R., | | | |
| Bueno,M.R., Camargo,A.A., Camargo,L.E., Carraro,D.M., Carer,H., | | | |
| Colauro,N.B., Colombo,C., Costa,F.F., Costa,M.C., Costa-Neto,C.M., | | | |
| Coutinho,L.L., Cristofani,M., Dias-Neto,E., Docena,C., El-Dorri,H., | | | |
| Faciocani,A.P., Ferreira,A.J., Ferreira,V.C., Ferro,J.A., | | | |
| Garra,J.S., Franca,S.C., Franco,M.C., Fromme,M., Furian,L.R., | | | |
| Gardner,M., Goldman,G.H., Goldman,M.H., Gomes,S.L., Gruber,A., | | | |
| Ho,P.L., Hobeisel,J.D., Junqueira,M.L., Kemper,E.L., Kitajima,J.P. | | | |
| and Marino,C.L. | | | |
| The genome sequence of the plant pathogen Xylella fastidiosa. The | | | |
| Xylella fastidiosa Consortium of the Organization for Nucleotide | | | |
| Sequencing and Analysis | | | |
| Nature 406 (6792), 151-157 (2000) | | | |
| 20365717 | | | |
| 10910347 | | | |
| 2 (bases 1 to 11541) | | | |
| Simpson,A.J.G., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M., | | | |
| Alvarenga,R., Alves,L.M.C., Araya,J.E., Bala,G.S., Baptista,C.S., | | | |
| Bartos,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.S., | | | |
| Bueno,M.R.P., Camargo,A.A., Camargo,L.E.A., Carraro,D.M., | | | |
| Carer,H., Colauro,N.B., Colombo,C., Costa,F.F., Costa,M.C.R., | | | |
| Costa-Neto,C.M., Coutinho,L.L., Cristofani,M., Dias-Neto,E., | | | |
| Docena,C., El-Dorri,H., Faciocani,A.P., Ferreira,A.J.S., | | | |
| Ferreira,V.C.A., Ferro,J.A., Fraga,J.S., Franca,S.C., Franco,M.C., | | | |
| Fromme,M., Furian,L.R., Gardner,M., Goldman,G.H., Goldman,M.H.S., | | | |
| Gomes,S.L., Gruber,A., Ho,P.L., Hobeisel,J.D., Junqueira,M.L., | | | |
| Kemper,E.L., Kitajima,J.P., Krieger,J.E., Kurmae,E.E., Laigret,F., | | | |
| Lambais,M.R., Leite,L.C.C., Lemos,E.G.M., Lemos,M.V.F., Lopes,S.A., | | | |
| Lopes,C.R., Machado,J.A., Machado,M.A., Madeira,A.M.B.N., | | | |
| Madeira,M.F., Martino,C.L., Marques,M.V., Martins,E.A.L., | | | |
| Martins,E.M.F., Matsukuma,A.Y., Menck,C.F.M., Miracca,E.C., | | | |
| Myaki,C.Y., Monteiro-Vitorello,C.B., Moon,D.H., Nagai,M.A., | | | |
| Nascimento,A.L.T.O., Netto,L.E.S., Nhani,Jr.,A., Nobrega,F.G., | | | |
| Nunes,L.R., Oliveira,M.A., de Oliveira,M.C., de Oliveira,R.C., | | | |
| Palmieri,D.A., Paris,A., Peixoto,B.R., Pereira,G.A.G., Pereira | | | |
| Jr.,H.A., Pesquero,J.B., Quaggio,R.B., Roberto,P.G., Rodrigues,V., | | | |
| de M. Rosa,A.J., de Rosa Jr.,V.E., de Sa,R.G., Santelli,R.V., | | | |
| Sawasaki,H.E., da Silva,A.C.R., da Silva,P.R., da Silva,A.M., Silva | | | |
| Jr.,W.A., da Silveira,J.F., Silvestri,M.L.Z., Siqueira,W.J., de | | | |
| Souza,A.A., de Souza,A.P., Terenzi,M.F., Truffi,D., Tsai,S.M., | | | |
| Tsubako,M.H., Vallada,H., Van Sluys,M.A., Verjovski-Almeida,S., | | | |
| Velho,A.L., Zago,M.A., Zatz,M., Meidanis,J. and Setubal,J.C. | | | |
| Direct Submission | | | |
| Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and | | | |
| Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP | | | |
| 13083-970, Brazil | | | |

| FEATURES | Source |
|--|--------|
| Location/Qualifiers | |
| 1..11541 | |
| /organism="Xylella fastidiosa 9a5c" | |
| /db_xref="taxon:160492" | |
| /clone="9a5c" | |
| complement(140..1057) | |
| /gene="Xf1258" | |
| complement(140..1057) | |
| /gene="Xf1258" | |
| /note="similar to SPIP11666 (percent identity: 35 %/query alignment coverage: 88.5 %/subject alignment coverage: 94.4 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder" | |
| /codon_start=1 | |
| /transl_table=1 | |
| /product="small conductance mechanosensitive ion channel" | |
| /protein_id="AAP84067.1" | |
| /db_xref="GI:9106243" | |
| /translation="MDWTKLISNIVSPATVTTTHGRPDWRIDMTHITMLNTAAAIATLIGVWLSKRLSNGLRLILRAHVEITLANELRNTAVALLIVSVSYLORIGVATSLAAVVGATGLAVGLAKDLSNIAAGVAILVGPVSGDHVIAOEGIDEIRIFOTRLRPFDORLITLPNSTITVPLINSTLPTRIELTVGVSYODNKOAOELLKARBNPKLIDDPALVYVNNIGESSVDLLILAVTODNNPNPAKSELBQIHNGLPENGINTIYQRLDILXHDNTNRKIASLIPVDAES" | |
| complement(1058..3457) | |
| /gene="Xf1259" | |
| complement(1058..3457) | |
| /gene="Xf1259" | |
| /note="similar to SPIP23538 (percent identity: 64 %/query alignment coverage: 99.9 %/subject alignment coverage: 100.8 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder" | |
| /codon_start=1 | |
| /transl_table=1 | |
| /product="phosphoenolpyruvate synthase" | |
| /protein_id="AAP84068.1" | |
| /db_xref="GI:9106244" | |
| /translation="MALNGASVYLNENILMHELRITDLARVGNSSIGENIGITLALGVSPGVATTAAPKAFIYANNLSKITRYKALLDVEDVKALPAAGKEIDWIIDPLOPELEKIDIRADYDTLSKNSGNDVAAVAEADLPDASFAGQOETFLNVGADQVHKIKIEVFASLYNDRAIYAVRHVGHEDVFSAIGIOLWVRSDISAGVLTDSGSGFRDVFVTASFGIGEMVVOGAVNPDEFFYKTKTKAGKPAILLRTGSKUYRVYSDSPGRVKIETDPTETLRHCFATSDVDVLELAKOALITEHNGRPMIDEMADGVSKEIVQARPEVTSRSHTTQIERPALQKREILTEGNAISQKSGISRVVHSIEDNHNOPGDVLVADMTDPMPEVPMKASAVTRNGRTCHAAITAEELGVAVGNGALSIIDGQPIVTSACGDEGYIYAGKLPETRTTDLGNHPPAPIKIMNVANPERAFDEGSLP NAGIGLARLEIILAAHIGVPRKALEYHKODADPKOKIDAKIAGIPRPSYRVRLAE GIATLTASVAPHTVIVRLSPKSNVYAHILGCGHPEOERPMKFGPGASRYNDPST DAPALEKAVIKVNNENGDMLRWYMPFVRLDEGRKIVIVLRNGLRGNGKLIVM MCEVFSNALADERLDFDGFSISNDLTLSLGDSSIVAFLEDERNPVAKKILS MAIKAAASKGRTYIGICGQPSDHPDLAEWMQEGIESLSPNPTVVDVTLAKIKTV H" | |
| 3549..4403 | |
| /gene="Xf1260" | |
| 3549..4403 | |
| /gene="Xf1260" | |
| /note="similar to SPI054458 (percent identity: 41 %/query alignment coverage: 100.7 %/subject alignment coverage: 98.6 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder" | |
| /codon_start=1 | |
| /transl_table=1 | |
| /product="conserved hypothetical protein" | |
| /protein_id="AAP84069.1" | |
| /db_xref="GI:9106245" | |
| /translation="MCRGSIIEGCMSTIRPVYSDGTGTAETIGHSITLTPSGFTFYTRKPTIDDAEKARDAQSRIETLAASQRIKRYRPVIVNSCPNFTSYLASSGALMDV FAPFTGLHEHLNLSRSRIGRAAGVDFEYHRIANAMFALAHGDGVAASDEADV ILVAVSRAGKPTCIYALAHYGRAANYPLIDEDLNSDOLPLRLRPYRKILFGITLIP ERLIOOIERPNSRYAALDPCKREVAAERMFASERITLSTHTSTIEIISVILVT LGIQREMF" | |
| 4501..4986 | |
| /gene="Xf1261" | |

gene

```

gene
5036 . 5542
/gene-"XF1262"
/note-"hypothetical protein: identified by sequence
similarity: putative; ORF located using Glimmer/RBSfinder"
/codon_start-1
/transl_table-1
/product-"hypothetical protein"
/protein_id-"AA084070.1"
/db_xref-"GI:9106246"
/translation-"MMLHQRIPRLSRLSWMLAALYENYRLKRLFAEDLEKGSYSVS
IGSDIDLCIDPLTIEHNRRTVDRLTYEELCDPMTGSPDSAAVRLYRDALQEAATCYGGR
RMDVDIGLSPLPGLMHHNRKMNIFLGKMLDYLAERGHGATLCAIGSRHVRVNRV
G"
5036 . 5542
/gene-"XF1262"
5036 . 5542
/gene-"XF1262"
/note-"similar to SPIP41354 (percent identity: 38 %/query
alignment coverage: 81.0 %/subject alignment coverage:
88.3 %); identified by sequence similarity: putative; ORF
located using Glimmer/RBSfinder"
/codon_start-1
/transl_table-1
/product-"7,8-dihydro-8-oxoguanine-5-triphosphatase"
/protein_id-"AA084071.1"
/db_xref-"GI:9106247"
/translation-"MLNVHAYTPPIATIGYVLSPPQCRVLTMRNARSQDVLHGRY
NGLCGKEPGEVDLAGIRRELYEAGIDCLISIRGRITSMFGKGGDMFGFEVVIDI
AFQETTHNAEGVLEWVERQRLFELPMMEDANFLVLPDDDRPHGVMPYCEGRM
OSWFSRL"
complement(5771 . 5965)
/gene-"XF1263"
complement(5771 . 5965)
/gene-"XF1263"
/note-"hypothetical protein: identified by sequence
similarity: putative; ORF located using Glimmer/RBSfinder"
/codon_start-1
/transl_table-1
/product-"hypothetical protein"
/protein_id-"AA084072.1"
/db_xref-"GI:9106248"
/translation-"MRISFPIHVPFPMHRLTAISSGMPSFNRIALLSIWKITTHNS
TOTLCPIPSQSHRAIOPENES"
complement(6020 . 6868)
/gene-"XF1264"
complement(6020 . 6868)
/gene-"XF1264"
/note-"similar to GI14377107 (percent identity: 23 %/query
alignment coverage: 66.0 %/subject alignment coverage:
27.4 %); identified by sequence similarity: putative; ORF
located using Glimmer/RBSfinder/Start codon shift: 780"
/codon_start-1
/transl_table-1
/product-"conserved hypothetical protein"
/protein_id-"AA084073.1"
/db_xref-"GI:9106249"
/translation-"MDLROGILTRLOHCOTIDAGOSCAVOTAISSIDGKRDVAGHIG
KICGFEPTESAGLSIDRSISRLPSDYLNKKGLAGLYAQMNTARHGSQWYTRVAV
AFNRYNTVQRPVLENTAEGICTSRMRGRGSLSEAGTFSNDGVLGMLHIGGTYRNV
SRATYSQDNAAPFTTGEASFKRTSAVLGADATIPLTTHLKMMAVVIDRLAKNRDAV
YSADANYIGNFNHRKIKGTATLSGLQYTISEKILLGLTLDVNOTAFGLIARGCTL
SVGGRF"
complement(7563 . 8492)
/gene-"XF1265"
complement(7563 . 8492)
/gene-"XF1265"
/note-"similar to GI14377108 (percent identity: 27 %/query
alignment coverage: 88.3 %/subject alignment coverage:
74.8 %); identified by sequence similarity: putative; ORF

```

| | | | |
|----|-------|---|-------|
| QY | 156 | cgccgcagcagcagctgcgaacatccctatgctggagagcgacactcttcgtcaaccgctacg | 215 |
| Db | 11171 | CGCTGAAAGCTCATATCTGCATTAACCCCTTTGTTGACGCAACCTCCTATCTGATATTCAGATTA | 11112 |
| QY | 216 | ggcgcaagaagtacagcgcgagcgcgagccagaccatgcccactctcgcgcgagaaat | 275 |
| Db | 11111 | TTWCAAAAGACATGATACCTTCATTCGGAAGCAACGAGATGTCTATTTAAAGAAAAAT | 11052 |
| QY | 276 | ggcgctcggttcccaatctcgaacgcgcgtctggatggacgcgatctgctgcatcaag | 335 |
| Db | 11051 | GCAAAACATTAAAGAAATTAACCTACAGAGAGTTTGCTTGATACATATGATGCAAGTAAAG | 10992 |
| QY | 336 | cgtaacagcgcgagccggc-----ggaagaaatactctgagccgccccctccca | 386 |
| Db | 10991 | TGCTCCCAAAAATTCAGAGAGCTCTTAATCTCAAGAGCCATCTAGACGCTGCAATTAACCA | 10932 |
| QY | 387 | gcagcagggaaaccacccctgaagtcattgattgattcattgcatctgcgtccggagcgga | 446 |
| Db | 10931 | GAAAAAAGGAATATCGGCTTAATACAGAACCTTTGGTATTTATAGACATACCTGAGCGCA | 10872 |
| QY | 447 | ctcgcgagcgctgcgtccctcaacggcggaactgcccgttaagcagcaggtttgcagacta | 506 |
| Db | 10871 | TTGCCAGCATTAACCTTCGATGATGTAATTAACCTTAACCCCGAAGAGCTCGCAACGTTA | 10812 |
| QY | 507 | tgaagcagtaacatgcgatctgagtagtaccctggagaaacggaagtaccacgcct | 566 |
| Db | 10811 | TAAAGAGATATATTTATACGATATGCGCATATTTTTTCAAAATCCAAAATATAAAGATAT | 10752 |
| QY | 567 | gcgagatcgltacgatactatgacgcgagctcgctcgccaaagcggtacacataatgacat | 626 |
| Db | 10751 | CCGATTTGTCAATATATTTAGACCGGATAGCTTGGCAAAATATATACCAACATGAGTAC | 10692 |
| QY | 627 | tcaagcgtgt-----gcaacgcggtgcccgtattacagcaagcatcagagtacgcct | 680 |
| Db | 10691 | TCCTAGTTGTCTCAACCAAAAAATACCGGAATTTACGAAGAGGATTAATAATATCCCT | 10632 |
| QY | 681 | cagaaattgcagccatctcgaagcgtgtatcatctatcatatgagcgccacactccgctgt | 740 |
| Db | 10631 | TAAATAACTTTCTGAATTTCCAAATGTATACAAATTACATGACATGCGCCACTCTGGCTG | 10572 |
| QY | 741 | gcttgggtggccataatgcccagcgatagctacaggaagtcacagaagtcctcaaa--- | 797 |
| Db | 10571 | GCTAGATGGGATACTAACCGGATACACAGACAGTCTTTGTACACCAAGTCAATTCAAG | 10512 |
| QY | 798 | cgcgagcatcggggtacacggcatcgagcgttcgttcccaacacagcggaattacacgc | 857 |
| Db | 10511 | CACATACAGCAGGTTTCTAGTGTCTATGTTTGGCACCAACACAGCAACACTCTACCC | 10452 |
| QY | 858 | gttaaggaagcgttcatgacgcgcaaccgc---agttcgggggagcagcggttgagtc | 914 |
| Db | 10451 | GTTATATGAACCAAACTACCAAAATCTGTATCTAAATATGGGGAGACGCTATTTAGATC | 10392 |
| QY | 915 | ggcgaattctaccagtggaatccctgacatcgacgaaagccgaatacgcgttgactgtta | 974 |
| Db | 10391 | CTCTAAATTCACGAATGGAATCGTATTTTGAAGAAATGATATTCAGAAACTTTGTGA | 10332 |
| QY | 975 | ctcgcggtcgtcgcgcggtgctttccaaagcagcatcgcatgctcatcgaaaccttacc | 1034 |
| Db | 10331 | TAAAGATTTTGTCTGCTGCTGCGCCATCCATATCGATATGATATGATACTGAGAG | 10272 |
| QY | 1035 | caagcgttgggtgtgtctcgaaagcaacaaagcccggaagcgcgacgcatgttacaac | 1094 |
| Db | 10271 | TAAAGGATGGGgtGACCTTGACGCCCCAAAGCTGCC---TTTGGAAAGAGATTTAATAG | 10215 |
| QY | 1095 | cttcgtcaaccagtcgaaagtattgaccttgcgacacacgcgcgcctgtgtgtcaaccagaa | 1154 |
| Db | 10214 | TTATGTGAATTCGAGCGTATAGACGCTGCTAATCAACCGTGAATGGGTAAATGAGAA | 10155 |
| QY | 1155 | cggtgtcggtcgtcgccagccgcgcgcaagcaccgcgaagacttccgaaagcgacct | 1214 |
| Db | 10154 | AGATGACAGGGATAGGTTTACTCCAAATTGCA-----CTCCAGAGTGCTCATGT | 10107 |

[illegible]

| Db | Accession | Gene | Strain | Source | Reference | Notes |
|------------|-----------|--|---|--------|-----------------|-------|
| Db | 1468 | GCCCTAGGCCAAGC--- | ggcAACCCCTCGGGGTGAGATACGGCGGGGCGACACGACGGCCAC | 1524 | | |
| Qy | 1800 | tgataaccagatcaaacccgggtctccagtttggtgaataaccgggtcgctcggtgagattt | 1859 | | | |
| Db | 1525 | CGACACACAGATGAAACCCGACAGCTCCGATCGTCAACACCCGGCTCGCAAGCCGTCCCGCT | 1584 | | | |
| Qy | 1860 | gtcgacgggtgacgggtgctgacatgtgttcaaccggagatggtgggttcgacacgtgtgta | 1919 | | | |
| Db | 1585 | GACCGACGCTGGAAGTGGCCCTACTGGTTACAC--- | GAAAGAACTCGACGACGACGGCCGAACAGTA | 1641 | | |
| Qy | 1920 | caactgtgacatggtggtcggtgagatggtgtgtggaataatccgagctcgttcgctcggtgaa | 1979 | | | |
| Db | 1642 | CTTCTGCGCACTGGGCGCGAGATCGCTGCTCGCAACATCGGGGCGGAGTTGCTGCTGTC | 1701 | | | |
| Qy | 1980 | ccggcgacgacgcgaaaggagggacactactactgaagtgtgtcttccatctgtgtggaacttggc | 2039 | | | |
| Db | 1702 | GCAGCCGGTCAAGCGGGGGGAGACGCTACATCAGCTGAGCTTCACGCGGGCGGAAGCATTC | 1761 | | | |
| Qy | 2040 | cgctgtgtgtgtgacgaggtgtgagatcaaacacgggtgtgaataagagtgacatgttcgaatt | 2099 | | | |
| Db | 1762 | GCGCGGAGGCAACACGCGGCGAGATTTCAGAACCGGATTCACCTTCACGAACTGGATGAAC | 1821 | | | |
| Qy | 2100 | caccgagacccaatgactactcgt--- | atggagacgaacacacccctccagactgtgacga | 2156 | | |
| Db | 1822 | CAACCAAAACGGGACGCTGCTGTCTACAAACGGGGCGGACGAGACGTTGGGGCCCTTCACCG | 1881 | | | |
| Qy | 2157 | ggtgacggtgtgacgtgacgagcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt | 2216 | | | |
| Db | 1882 | GATTTCGCTTTATTCGCAACGGGCTCTGTGTATAGGGGGAGAGCCGGCGCGGATTCGTC | 1941 | | | |
| Qy | 2217 | cagcccaacacacatcccgagcccgagcccgagcccgagcccgagcccgaggtgtgtgtgtgtgt | 2276 | | | |
| Db | 1942 | GCCCGCGAGCGCGGAGGTCACACCGACACCGACCGACCGGCGAGTCACACCGCACACC | 2001 | | | |
| Qy | 2277 | gccgagtggtgcgcac | 2291 | | | |
| Db | 2002 | TACGCTACGCCAAC | 2016 | | | |
| RESULT | 15 | | | | | |
| AF078737 | | | | | | |
| LOCUS | | 11707 bp | DNA | linear | BCT 31-JAN-2001 | |
| DEFINITION | | Caldicellullosiruptor sp. Tok7B.1 xyna (xyna) and celB (celB) genes. | | | | |
| ACCESSION | | AF078737 | | | | |
| VERSION | | AF078737.1 | | | | |
| KEYWORDS | | GI:4836163 | | | | |
| SOURCE | | | | | | |
| ORGANISM | | Caldicellullosiruptor sp. Tok7B.1. | | | | |
| REFERENCE | | Caldicellullosiruptor sp. Tok7B.1 | | | | |
| AUTHORS | | Bacteria: Firmicutes; Bacillus/Clostridium group; Syntrophomonadaceae; Caldicellullosiruptor. | | | | |
| TITLE | | 1 (bases 1 to 11707) | | | | |
| JOURNAL | | Gibbs, M.D., Reeves, R.A., Farrington, G.K., Anderson, P., Williams, D.P. and Bergquist, P.L. | | | | |
| MEDLINE | | Multidomain and multifunctional glycosyl hydrolases from the extreme thermophile Caldicellullosiruptor isolate Tok7B.1 | | | | |
| PUBMED | | Curr. Microbiol. 40 (5), 333-340 (2000) | | | | |
| REFERENCE | | 2 (bases 1 to 11707) | | | | |
| AUTHORS | | Gibbs, M.D., Reeves, R.A., Farrington, K.G., Anderson, P., Williams, D.P. and Bergquist, P.L. | | | | |
| TITLE | | Direct Submission | | | | |
| JOURNAL | | Submitted (17-JUL-1998) School of Biological Sciences, Macquarie University, Sydney, New South Wales 2109, Australia | | | | |
| FEATURES | | Location/Qualifiers | | | | |
| SOURCE | | 1..11707 | | | | |
| CDS | | /organism="Caldicellullosiruptor sp. Tok7B.1" | | | | |
| | | /strain="Tok7B.1" | | | | |
| | | /db_xref="taxon:80339" | | | | |
| | | <1..683 | | | | |
| | | /note="Orf1" | | | | |

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 31, 2002, 09:56:38 ; Search time 378.37 Seconds

(without alignments)
16730.353 Million cell updates/sec

Title: US-09-917-384-2

Perfect score: 3687

Sequence: 1 atggagcgcaaccacaatc.....tcagctgcacagccagctga 3687

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues 3472872

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08

Maximum Match 1008

Listing first 45 summaries

Database : N.Geneseq_032802:*

1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*

2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*

3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*

4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*

5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*

6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*

7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*

8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*

9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*

10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*

11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*

12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*

13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*

14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*

15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*

16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*

17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*

18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*

19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*

20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*

21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*

22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*

23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*

24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|----------|--------------------|
| 1 | 423.6 | 11.6 | 3004 | 17 | AA12337 | A. cellulolyticus |
| 2 | 423.6 | 11.5 | 3004 | 21 | AA255924 | Acidothermus cellu |
| 3 | 324 | 8.8 | 3503 | 19 | AAV07164 | Thermomonospora fu |
| 4 | 322.4 | 8.7 | 3404 | 17 | AA15595 | Thermomonospora fu |
| 5 | 314.6 | 8.5 | 1269 | 19 | AA15596 | Thermomonospora fu |
| 6 | 251.2 | 6.8 | 3033 | 19 | AA194197 | Thermomonospora fu |
| 7 | 134.2 | 3.6 | 11706 | 20 | AA55661 | DNA sequence encod |
| 8 | 134.2 | 3.6 | 11707 | 24 | AA26525 | Active cellulase p |
| 9 | 116.4 | 3.2 | 2977 | 12 | AA013001 | Endo1 gene encodin |

| | | | | | | |
|----|-------|-----|---------|----|-----------|--------------------|
| 10 | 108.4 | 2.9 | 1422 | 20 | AA27948 | H. insolens Cel6A |
| 11 | 108.4 | 2.9 | 1437 | 20 | AA27958 | H. insolens Cel6A |
| 12 | 98.8 | 2.7 | 6415 | 20 | AA55662 | DNA sequence encod |
| 13 | 98.8 | 2.7 | 6416 | 24 | AA26526 | Active cellulase p |
| 14 | 97.6 | 2.6 | 1621 | 19 | AA07163 | Thermomonospora fu |
| 15 | 96.2 | 2.6 | 6510 | 22 | AA54668 | Tumour suppressor |
| 16 | 96.2 | 2.6 | 6510 | 24 | ABJ2395 | Human immune syste |
| 17 | 96.2 | 2.6 | 6510 | 24 | AA61115 | Human gene regulat |
| 18 | 95.2 | 2.6 | 1116 | 20 | AAV72359 | Actinomyces sp. 3 |
| 19 | 95.2 | 2.6 | 1175 | 21 | AA57031 | Actinomyces cell |
| 20 | 95.2 | 2.6 | 1715 | 21 | AA57031 | DNA sequence of an |
| 21 | 95.2 | 2.6 | 2942 | 21 | AA48661 | L. mycopilius chl |
| 22 | 94.8 | 2.6 | 390 | 13 | AAQ21833 | Randomising oligon |
| 23 | 94.8 | 2.6 | 390 | 14 | AAQ36859 | PCR primer for 5' |
| 24 | 94.8 | 2.6 | 390 | 22 | AA76910 | Sequence containin |
| 25 | 88.6 | 2.4 | 4376 | 21 | AA14818 | DNA encoding a Cl |
| 26 | 87.8 | 2.4 | 390 | 13 | AAQ21833 | Randomising oligon |
| 27 | 87.8 | 2.4 | 390 | 14 | AAQ36859 | PCR primer for 5' |
| 28 | 87.8 | 2.4 | 390 | 22 | AA76910 | Sequence containin |
| 29 | 87.8 | 2.4 | 4403765 | 22 | AA199683 | Mycobacterium tube |
| 30 | 87.6 | 2.4 | 37286 | 23 | AA59522 | Humicola insolens |
| 31 | 86.2 | 2.3 | 2409 | 19 | AAV19377 | H. insolens cellul |
| 32 | 86.2 | 2.3 | 2409 | 19 | AAV19281 | Portion of PAEC-1 |
| 33 | 84.2 | 2.3 | 2600 | 12 | AAQ15178 | Cellulohydrolase |
| 34 | 83.6 | 2.3 | 1683 | 6 | AAAN50359 | Trichoderma reesei |
| 35 | 82.6 | 2.2 | 3241 | 21 | AAAF4880 | Streptomyces nous |
| 36 | 82.6 | 2.2 | 65140 | 22 | AA17184 | Streptomyces nous |
| 37 | 82.6 | 2.2 | 125401 | 22 | AA17186 | Streptomyces nous |
| 38 | 81.6 | 2.2 | 4403765 | 22 | AA199683 | Mycobacterium tube |
| 39 | 81.2 | 2.2 | 2029 | 20 | AA55660 | DNA sequence encod |
| 40 | 81.2 | 2.2 | 2029 | 24 | AA26568 | Active cellulase h |
| 41 | 81 | 2.2 | 4411529 | 22 | AA19582 | Mycobacterium tube |
| 42 | 78.2 | 2.1 | 4056 | 22 | AAH75060 | Nucleotide sequenc |
| 43 | 77.6 | 2.1 | 2712 | 20 | AA06825 | Streptomyces albid |
| 44 | 77.6 | 2.1 | 2712 | 22 | AAH74538 | Nucleotide sequenc |
| 45 | 77.4 | 2.1 | 985 | 19 | AAV64348 | M. tuberculosis Im |

ALIGNMENTS

| | |
|----------|---|
| RESULT 1 | |
| AA12337 | AA12337 standard; DNA: 3004 BP. |
| ID | AA12337 |
| XX | AA12337; |
| AC | 08-OCT-1996 (first entry) |
| XX | |
| XX | A. cellulolyticus EI endoglucanase gene. |
| DE | |
| XX | EI endoglucanase; cellulase; cellulose; saccharification; ethanol; |
| KW | ss. |
| XX | |
| OS | Acidothermus cellulolyticus. |
| XX | |
| FH | key |
| FT | misc_difference 200 |
| FT | location/Qualifiers |
| FT | /*tag- a |
| FT | /note- "base g at position 200 is shown in Fig 1, but is missing in the sequence given on page 22 of the specification" |
| FT | 710..725 |
| FT | /*tag- b |
| FT | /note- "putative upstream regulatory sequence" |
| FT | 772..779 |
| FT | /*tag- c |
| FT | 781..790 |
| FT | /*tag- d |
| FT | /rpt_type- DIRECT |
| FT | 795..804 |
| FT | /*tag- e |
| FT | /rpt_type- DIRECT |

```

FT repeat_unit      810..817   /*tag= f
FT                                     /rpl_type= DIRECT
FT CDS              824..2512
FT                                     /*tag= g
FT sig_peptide     824..946
FT                                     /*tag= h
FT sig_peptide     863..946
FT                                     /*tag= i
FT mat_peptide     /note= "alternative signal sequence"
FT                                     947..2509
FT                                     /*tag= j
FT terminator      2514..2560
FT                                     /*tag= k
XX
XX MO9602551-A1.
PD 01-FEB-1996.
XX
XX 14-JUL-1995;    95WO-US08868.
XX
PR 15-JUL-1994;    94US-0276213.
PR 26-SEP-1989;    89US-0412434.
PR 27-JAN-1992;    92US-0826089.
PR 21-SEP-1993;    93US-0125115.
XX
PA (MIDE ) MIDWEST RES INST.
XX
PI Himmel ME, Laymon RA, Thomas SR.
XX
XX WPI; 1996-105843/11.
DR P-PSDB; AAR89927.
XX
PT New isolated DNA encoding endoglucanase - obtd from Acidothermus
PT cellulolyticus, used for prodn of the enzyme for use in cellulose
PT hydrolysis.
XX
PS Claim 4; Page 22-24; 34pp; English.
XX
CC The E1 endoglucanase gene (T12337) of Acidothermus cellulolyticus
CC codes for an enzyme, the mature form of which (R89927) can hydrolyse
CC cellulose and xylan. The gene was obtd. by ligating Sau3A fragments
CC of A. cellulolyticus genomic DNA into EMBL3 lambda phage arms,
CC packaging the phage DNA, plating on E.coli LE392 agar contg. CMC,
CC staining with Congo Red, and isolating DNA from positive clones.
CC Alternative signal sequences were identified (see also R89928-29).
CC The gene can be incorporated into a vector and used for large-scale,
CC low-cost prodn. of recombinant E1 endoglucanase, pref. using
CC Saccharomyces, Zyomonas or E.coli hosts. The enzyme is useful
CC for the saccharification of cellulosic biomass for fermentation to
CC ethanol.
XX
SQ Sequence 3004 BP; 604 A; 931 C; 938 G; 531 T; 0 other:
XX
Query Match          11.6%; Score 427; DB 17; Length 3004;
Best Local Similarity 99.8%; Pred. No. 2,4e-65;
Matches 438; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
OY 1 atggagcggaaccacaacatccgagcgcgttcaggatctccagagaggaagcacagaatg 60
Db 2567 atggagcggaaccacaacaaacgcgagcgcgttcaggatctccagagaggaagcacagaatg 2626
OY 61 ccgcgcattctcaaaaacgagctgcgaacgcgcgttcctcgcgcgcggcggttgatcgacgcc 120
Db 2627 ccgcgcattctcaaaaacgagctgcgaacgcgcgttcctcgcgcgcggcggttgatcgacgcc 2686
OY 121 tccatctgctgcgcttgcgatgcagcatctctgcattcgcgcgcgcgaacgctgcacaatccc 180
Db 2687 tccatctgctgcgcttgcgatgcagcatctctgcattcgcgcgcgcgaacgctgcacaatccc 2746
OY 161 tatgggggagcgactctttctgttaaccctgctcgtggcgccaagaagttaaagcgcaagcg 240

```

| | | | |
|----------|--|--|------|
| Db | 2747 | tatcgcgagcgcgcctcttcctgcaccccgctactcgtgcgccgaagatcagagcgaa-cg | 2805 |
| Qy | 241 | gcgaaccagacccaatgccaactctcgcgcagcgaatacgcgctggttccacatctgcagc | 300 |
| Db | 2806 | gcgaccacagaccgaatgccactctcgcagcgaataatgcgctggttccacatctgcagc | 2865 |
| Qy | 301 | gcgcctcgatgtagccgcgcacatcgtctgcatcaacgcgcgtcaacgcgcgcgcgttcagc | 360 |
| Db | 2866 | gcgcctcgatgtagccgcgcacatcgcctcgcacatcgcacgcgcgtcaacgcgcgcgcgttcagc | 2925 |
| Qy | 361 | acatatctgacgcgcgcgcctctccacgcagcagcgaacccaccctgaagtcattgaagt | 420 |
| Db | 2926 | acatatctgacgcgcgcgcctctccacgcagcagcgaacccaccctgaagtcattgaagt | 2985 |
| Qy | 421 | gtcatctcagcatctgcgcg | 439 |
| Db | 2986 | gtcatctcagcatctgcgcg | 3004 |
| RESULT 2 | | | |
| ID | AA255924 | standard; DNA; 3004 BP. | |
| AC | AA255924; | | |
| DT | 10-APR-2000 | (first entry) | |
| DE | Acidothermus cellulolyticus EI endoglucanase gene. | | |
| KW | EI endoglucanase; cellulose binding domain; CBD; cellulose modification; beta-1,4-endoglucanase; endocellulase; thermostable; ds. | | |
| OS | Acidothermus cellulolyticus. | | |
| XX | Key | Location/Qualifiers | |
| FT | CDS | 824..2512 | |
| FT | | /tag- a | |
| FT | sig_peptide | 824..946 | |
| FT | | /tag- b | |
| FT | sig_peptide | 863..946 | |
| FT | | /note- "Encodes putative signal peptide" | |
| FT | mat_peptide | 947..2509 | |
| FT | | /tag- c | |
| FT | | /note- "Encodes putative signal peptide (alternative)" | |
| FT | | /tag- e | |
| FT | misc_feature | 2204..2509 | |
| FT | | /product- "Mature EI endoglucanase" | |
| FT | | /tag- f | |
| FT | | /note- "Encodes cellulose binding domain (specifically claimed)" | |
| XX | CA2226898-A1. | | |
| XX | 25-SEP-1999. | | |
| XX | 25-MAR-1998; | 98CA-2226898. | |
| XX | 25-MAR-1998; | 98CA-2226898. | |
| XX | (MIDE) MIDWEST RES INST. | | |
| XX | Laymon RA, Adney WS, Thomas SR, Himmel ME; | | |
| XX | WPI: 2000-087663/08. | | |
| XX | P-PSDB: AAY69508. | | |
| XX | Isolated domains of Acidothermus cellulolyticus EI endoglucanase useful for labeling or modifying a cellulose and for purifying or immobilizing a binding domain fusion protein to cellulose | | |
| XX | Claim 12; Fig 1; 85pp; English. | | |
| XX | This sequence represents the Acidothermus cellulolyticus EI endoglucanase | | |

[illegible]

| | | |
|----|--|---|
| AC | AA155595; | |
| XX | | |
| DT | 06-APR-1996 | (first entry) |
| XX | | |
| DE | Thermostable cellulase-E3 gene. | |
| XX | | |
| KW | Cellulase-E3; Thermomonospora fusca; thermostable; plasmid pS24; | |
| KW | plasmid pS26; Escherichia coli; cloning; vector; primer extension; | |
| KW | 5'-untranslated region; inverted repeat; Streptomyces lividans; | |
| KW | papain; catalytic domain; cellulose; chitosan; protease resistance; | |
| KW | synergism; cellobiohydrolase; beta-glucosidase; saccharification; | |
| KW | surfactant additive; paper recycling; delinking; paper refining; ds. | |
| XX | | |
| OS | Thermomonospora fusca strain YX36. | |
| XX | | |
| FH | Key | Location/Qualifiers |
| FT | 5'UTR | 162..474 |
| FT | | /*tag= a |
| FT | repeat_unit | 250..263 |
| FT | | /*tag= b |
| FT | | /rpt_type= INVERTED |
| FT | | /note= "Binding site for cellobiose induction |
| FT | | regulatory protein" |
| FT | sig_peptide | 475..588 |
| FT | | /*tag= C |
| FT | CDS | 475..2265 |
| FT | | /*tag= d |
| FT | | /product= Cellulase-E3 |
| FT | mat_peptide | 589..2262 |
| FT | | /*tag= e |
| FT | misc_feature | 994..2043 |
| FT | | /*tag= f |
| FT | | /product= Catalytically active polypeptide (AAR90716) |
| FT | | /note= "Corresponds to AA15596" |
| FT | terminator | 2349..2366 |
| FT | | /*tag= g |
| FT | | /note= "Contains 14-base palindrome" |
| FT | stem_loop | 2349..2366 |
| FT | | /*tag= h |
| FT | stem_loop | 2368..2409 |
| FT | | /*tag= i |
| XX | | |
| PN | WO9600281-A1. | |
| XX | | |
| PD | 04-JAN-1996. | |
| XX | | |
| PF | 23-JUN-1995; | 95WO-US09069. |
| XX | | |
| XX | 24-JUN-1994; | 94US-0265429. |
| XX | | |
| PA | (CORR) CORNELL RES FOUND INC. | |
| XX | | |
| PI | Walker LP, Wilson DB, Zhang S; | |
| XX | | |
| DR | WPI; 1996-068865/07. | |
| DR | P-PSDB; AAR90715. | |
| XX | | |
| PT | Cellulase from Thermomonospora fusca - is thermostable and is useful | |
| PT | in a variety of industrial applications e.g. clarification of fruit | |
| XX | juices, fabric softening, etc | |
| PS | Claim 1; Page 32-37; 53pp; English. | |
| XX | | |
| CC | The sequence encodes thermostable cellulase-E3 from Thermomonospora | |
| CC | fusca, and has been isolated from a genomic NotI DNA library in | |
| CC | Escherichia coli (in plasmid pS24 and plasmid pS26). The | |
| CC | transcriptional start site has been determined using primer AA15597, | |
| CC | complementary to a region encoding the E3 signal peptide, and is | |
| CC | located 313 bases upstream of the translation initiation codon. The | |
| CC | long 5'-untranslated sequence contains putative regulatory sequences | |
| CC | and potential secondary structure. Preceding the 5'-mRNA start site | |
| CC | is a 13-bp sequence similar to a later 14-bp inverted repeat, which | |

CC may be involved in regulation of expression. The gene may be cloned and expressed in *E. coli* or *Streptomyces* 1141dans. The gene product, or a catalytically active polypeptide obtained by pepsin digestion, hydrolyses cellulose or chitosan at pH 5-11 and 40-70 °C, and has significant activity at over 60 deg C. E3 has good thermal stability to proteolysis in culture supernatant than T. fusca cellulases-E2 and -E5, and shows strong synergistic activity when combined with other cellulases, cellobiohydrolase and CC optionally beta-glucosidase. The cellulase and mixtures may be used e.g. in cellulose saccharification for ethanol production, fruit juice clarification, as a surfactant additive, or in delinking or refining of recycled paper.

SQ Sequence 3404 BP; 619 A; 1239 C; 1004 G; 542 T; 0 other;

| | | | | | | | |
|---------------------------|--------|------------|----------|--------|--------|-------|----|
| Query Match | 8.78; | Score | 322.4; | DB 17; | Length | 3404; | |
| Best Local Similarly | 56.58; | Pred. No. | 3.2e-47; | | | | |
| Matches 774; Conservative | 0; | Mismatches | 536; | Indels | 60; | Gaps | 7. |

[illegible]

| | | | |
|--|--|--|------|
| Dd | 1784 | acgagctctcccttgcgtccaggacctcgttcagggcccctgaatcgccaaagggtctccggtccg | 1843 |
| Oy | 1007 | gcatcgcatgctatcatcgacaacttcgcgaacggttggtgggtgtgtcgcagaacaaacaaag | 1066 |
| Dd | 1844 | acatcgtgatgtcatatcgtcaaacctccccgcgaacggctcgtgggtggcccgaaaccgtccgacg | 1903 |
| Oy | 1067 | gccgcgaacacccgcgcacgcatgtgtcaaaccttcgtccaaccagctcgaagattgaacttcggc | 1126 |
| Dd | 1904 | gaccagatcctctcacaccgaccttaacaacctagttgaacgaagagccgtatcgcagccgcta | 1963 |
| Oy | 1127 | agcacgcgcgcctgtgtgtgtcaaccagaacgggtggcggtcctcgtgccagccgcgcagcga | 1186 |
| Dd | 1964 | tccacccccgggttaaactgtgtcaaaccaagcccggtgtcggtcctcgtgcgaagc----- | 2011 |
| Oy | 1187 | gccgcgaagacttcccgaacgcgcacctcgtacgcgtatgtctcgtgaatcaagccgcgcgggt | 1246 |
| Dd | 2012 | ggcccaacggttcaaacccgcgcctcccggtgtgtgaacgctcagctcgtggatgaaagcccccggg | 2071 |
| Oy | 1247 | agtctgacggagacaagaagcgtcgcgcagcagctcgcacaactgtgcaagaagctcgaaacccaagt | 1306 |
| Dd | 2072 | agtcgcgaacggcgcagcagagagatcccgaaacgaaggaagggtcttcgcgcgcatagt | 2131 |
| Oy | 1307 | ggcagccgaagtlacagaacgctcgtacgggtgactgtaccaac-----gcgttacga | 1357 |
| Dd | 2132 | ggcagccgaactaacagaagggcaaaagccgcgcgaacgcacaaacccctcgggtgtgcgtgccca | 2191 |
| Oy | 1358 | actcccgcatcgcgcgcgaatgtgttctccggcggcgaagttgaaccaggttgtcgtgaaagcac | 1417 |
| Dd | 2192 | acgcgcccatctccgcgcgacctgtgtctctgcgcgaattccgcgagctgtcgtgccaacgctt | 2251 |
| Oy | 1418 | ggccagcgggtgcgaacgctcgcaccagctcgaagccgcgcgcctccgcgcgcgcgaagtcgctcg | 1477 |
| Dd | 2252 | accgcgcctctgtaaagcgggtgtgaaggcaacgcgtgcgaagccctcaacggaggaactgtatcg | 2311 |
| Oy | 1478 | cttcgcgcagatcgcagccgcgaatcgcgaagcccgaaagcagctcgcacatcgccg | 1527 |
| Dd | 2312 | cacctctctagcgcgagagacgcgcgcgtcaccatccctgtggcgggcgcg | 2361 |
| RESULT 5 AAT15596 standard; DNA: 1269 BP. | | | |
| ID | AAT15596 | standard; DNA: 1269 BP. | |
| AC | AAT15596: | | |
| XX | 06-APR-1996 | (first entry) | |
| DT | Thermostable cellulase-E3 catalytic domain gene. | | |
| DE | | | |
| XX | Cellulase-E3: Thermomonospora fusca; thermostable; papain; | | |
| KW | catalytic domain; Escherichia coli; cloning; Streptomyces lividans; | | |
| KM | cellulose; chitinase; protease resistance; synergism; | | |
| KV | cellobiohydrolase; beta-glucosidase; saccharification; | | |
| XX | surfactant additive; paper recycling; delinking; paper refining; ds. | | |
| XX | | | |
| OS | Thermomonospora fusca strain YK36. | | |
| XX | | | |
| PN | W09600281-A1. | | |
| XX | | | |
| PD | 04-JAN-1996. | | |
| XX | | | |
| PF | 23-JUN-1995; 95WO-US09069. | | |
| XX | | | |
| PR | 24-JUN-1994; 94US-0265429. | | |
| XX | | | |
| PA | (CORR) CORNELL RES FOUND INC. | | |
| XX | | | |
| PI | Walker LP, Wilson DB, Zhang S; | | |
| XX | | | |
| DR | WPI; 1996-068865/07. | | |
| XX | P-PSDB; AAR90716. | | |


```

AAK27948
ID   AAK27948 standard; DNA; 1422 BP.
XX
XX AC
XX AAK27948;
XX
XX DT
08-JUN-1999 (first entry)
XX
DE H. insolens Cel6A fungal cellulase coding sequence.
XX
XX Cel6B; Cel6A; fungal cellulase; cleaning composition; conditioner;
XX cellulolytically active protein; endo-1,4-beta-glucanase; enzyme;
XX fabric softener; Humicola-like cellulase; glycosyl hydrolase family 6;
XX detergent composition; ss.
XX
XX Humicola insolens.
OS
PN WO9901544-A1.
XX
PD 14-JAN-1999.
XX
PF 02-JUL-1998; 98WO-DK00299.
XX
PR 04-JUL-1997; 97DK-0000813.
XX
PA (NOVO ) NOVO-NORDISK AS.
PI Andersen KV, Damgaard B, Lund H, Nielsen JB, Schulein M;
DR WPI: 1999-106046/09.
P-PSDB; AAU01077.
XX
PT Cleaning composition containing Humicola endo-beta-1,4-glucanase -
PT useful as detergent compositions or additives, or as fabric
PT conditioners
PS
PS Example 3; Page 241-243; 271pp; English.
XX
XX This sequence encodes the Humicola insolens Cel6A fungal cellulase.
CC The invention relates to a cleaning composition (A) that contains at
CC least one enzyme (I) with cellulolytic activity and at least 25 wt.% of
CC total cellulolytically active protein present is Humicola
CC endo-1,4-beta-glucanase or a Humicola-like cellulase of the glycosyl
CC hydrolase family 6. (A) are used as detergent compositions or additives,
CC or as fabric softener or conditioner. (I) provides colour clarification,
CC and possibly soil removal, without significant weakening of cellulosic
CC materials during pre-soaking or wet storage (contrast use of enzymes from
CC families other than 6).
CC
SQ Sequence 1422 BP; 261 A; 529 C; 370 G; 262 T; 0 other;

Query Match          2.9%; Score 108.4; DB 20; Length 1422;
Best Local Similarity 49.9%; Pred No. 3.8e-10;
Matches 367; Conservative 0; Mismatches 351; Indels 18; Gaps 3

147 tcttcgcatcgccgcgcagcagcagctgcacatccctaigtgcggaggacacttcgtcaa 206
Db      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
315 tgcctgaccacttgccagctacaagcgcaacccttcggagggttcagtgttgggccaa 374
QY      cccgtacctggcgcaagaagtacagagcggaagcgcgcaaccagcaaatgccaacttcgc 266
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
375 caactactccggtccgaagttcacactcttcgcatccctcagatactgaccct----- 429
QY      agcgaaatgcgcgltcgtttccacatatccgacgcgcgtctgatgacgcgatgcgtgc 326
Db      -gccttgagggtcgtcgccctcgtctgctgctgcgaaggttcccagcttcagtgcgcgcg 488
QY      gatcaacggcgticaaacggcggaaccggcgttcgacatatctggaacgcgcgccctccca 386
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
489 ggaagcgtacgcgtcgcaaccctcgtctccagagaccctcttgatccggaggcgacaaa 548
QY      gcagcagggaaccaaccctgaagtaattgataattgcatactacgatactgcggagcga 446
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

| | | | |
|-----------|--|--|------|
| Db | 549 | agcgagcgcgaaatccccaataatgctgccaaatcgtcgttaccgacttgcctgaacgcga | 608 |
| Qy | 447 | ctgcgcgcgcgtctgcctccccaacgcgcgaactgcccgcgttaacgcgcagcgtttcgaaccta | 506 |
| Db | 609 | ctgcgcgtcgcgcgcgtcttcgcgaacgcgcgcgcgtggtggccacgcgaacaacgcgcgtcaacaacta | 668 |
| Qy | 507 | tgaacgcgcgtacatcgatccgcgaattgcgcgagtatccctgagaaatcgcgaagtaactcgaacct | 566 |
| Db | 669 | caagcatalacataacgcgaatccgcgagattcattcc-----ttctcgaagt | 719 |
| Qy | 567 | gcgcgtctgtgcgcgtatcttgagccgcgcgtcgtcgcgaacgcgcgttacaataatagacat | 626 |
| Db | 720 | ccgcacccattctgttcattgagccgcgaactcgcctgcgcgaacatgttaccacaatgtaacgt | 779 |
| Qy | 627 | tcaacgcgtgtcaacgcgcgcgtgtacttaccgaaagcgcgcgatgcgagtaacgcgtcaacga | 686 |
| Db | 780 | tcccaagtcgcgcgcgtgtgcgcgtctgcgaactccgcgcggttgaccattatgcctcgaacga | 839 |
| Qy | 687 | attgcacgcgaatccgcgaacgltgtacatactacatgtaacgcgcgcgaactccgcgtctgttg | 746 |
| Db | 840 | --gtcgcgaactccgcgaacgltgcgcgaatgtacatgtaacgcgcgcgaacgctgctgtcttg | 896 |
| Qy | 747 | gtgcgcgcgaataatgtccgcgcgcgtatcgtacacgcgcgaagtcacgaagttcttcaacgcgcgat | 806 |
| Db | 897 | ctgcgcgcgcgaacatccgaacgcgcgcgtgcgtcttctgcgcgaagatctacgaagatgcgcg | 956 |
| Qy | 807 | cgcgcgtcaacgcgcgtacgcgcgtctgcgcgaacgcgcgcgaattacgcgcgttgaaaga | 866 |
| Db | 957 | caagcccccgcgcgtctgcgcgtctgcgcgaacgcgcgaatcaataacgcgcgtgagcgt | 1016 |
| Qy | 867 | gccgttcattgacgcgc 882 | |
| Db | 1017 | ctgcgcgcgcgcgcgc 1032 | |
| RESULT 11 | | | |
| ID | AAx27958 | AAx27958 standard; DNA; 1437 BP. | |
| AC | AAx27958; | | |
| DT | 08-JUN-1999 | (first entry) | |
| DE | H. insolens Cel6A fungal cellulase coding sequence. | | |
| KM | Cel6B: Cel6A; fungal cellulase; cleaning composition; conditioner; cellulolytically active protein; endo-1,4-beta-glucanase; enzyme; fabric softener; Humicola-like cellulase; glycosyl hydrolase family 6; detergent composition; ss. | | |
| OS | Humicola insolens. | | |
| PN | WO9901544-A1. | | |
| PD | 14-JAN-1999. | | |
| PF | 02-JUL-1998; | 98WO-DK00299. | |
| PR | 04-JUL-1997; | 97DK-0000813. | |
| PA | (NOVO) NOVO-NORDISK AS. | | |
| PI | Andersen KV, Damgaard B, Lund H, Nielsen JB, Schulein M; | | |
| DR | WPI: 1999-106046/09. | | |
| DR | P-PSDB; AAY01077#. | | |
| PT | Cleaning composition containing Humicola endo-beta-1,4-glucanase | | |
| PT | useful as detergent compositions or additives, or as fabric | | |
| PT | conditioners | | |
| PS | Example 3; Fig 2; 271pp; English. | | |
| XX | | | |

This sequence encodes the Humicola insolens Cel6A fungal cellulase. The invention relates to a cleaning composition (A) that contains at least one enzyme (I) with cellulolytic activity and at least 25 wt. % of total cellulosylatically active protein present is Humicola endo-1,4-beta-glucanase or a Humicola-like cellulase of the glycosyl hydrolase family 6. (A) are used as detergent compositions or additives, or as fabric softener or conditioner. (I) provides colour clarification, and possibly soil removal, without significant weakening of cellulosic materials during pre-soaking or wet storage (contrast use of enzymes from families other than 6).

SQ Sequence 1437 BP; 266 A; 532 C; 374 G; 265 T; 0 other;

| | | | | |
|-----------------------|----------------|-------------------|-----------|-------------|
| Query Match | 2.98 | Score 108.4 | DB 20 | Length 1437 |
| Best Local Similarity | 45.98 | Pred. No. 3.8e-10 | | |
| Matches 367 | Conservative 0 | Mismatches 351 | Indels 18 | Gaps 3 |

[illegible]

RESULT 12
AA55662

ID AAX55662 standard; DNA; 6415 BP.

AC AAX55662;

DT 30-JUL-1999 (first entry)

DE DNA sequence encoding truncated cellulases.

KW Cellulase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;

KW cotton-containing fabric; stonewashing; ss.

OS Unidentified.

PN EP921188-A2.

PD 09-JUN-1999

15-SEP-1998; 98EP-0810919.

PR 19-SEP-1997; 97US-0932571.

PA (CLRN) CLARIANT FINANCE BVI LTD.

PI Anderson P, Bergquist PL, Daniels RM, Farrington GK;

| XX | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 | 71 | 72 | 73 | 74 | 75 | 76 | 77 | 78 | 79 | 80 | 81 | 82 | 83 | 84 | 85 | 86 | 87 | 88 | 89 | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 | 100 |
|----|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|-----|
| XX | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 | 71 | 72 | 73 | 74 | 75 | 76 | 77 | 78 | 79 | 80 | 81 | 82 | 83 | 84 | 85 | 86 | 87 | 88 | 89 | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 | 100 |

DR P-PSDB; AAY13493.

PT New truncated cellu

XX

XX

CC proteases of native thermophilic and

CC One or several sequences, e.g. Cel B/3, Cel E/1, Cel E1/2, Cel E1/2/3,
CC Cel E/6 or Cel E3/B/5, or a stability region from one of the defined full-
CC length sequences, or functional equivalents. Cel B/5 extends from amino
CC acid A1011 to P1424 or K1425 or N1426, and Cel B/4/5 extends from amino
CC acid R635 to N1426 in the sequence shown in AA133497; Cel E1 extends
CC from amino acid Y39 to D481. Cel E1/2 extends from Y39 to G633. Cel
CC E1/2/3 extends from Y39 to G812. Cel E/6 extends from amino acid V1233 to
CC K1751 and the stability region extends from amino acid E482 to G635 in
CC the sequence shown in AA13493; Cel E3/B/5 is shown in AA13494. The new
CC enzymes are useful in laundry detergent compositions to prevent or
CC remove staining, backstaining or graying, for use on cellulosic
CC materials including cotton-containing fabrics. They are especially useful
CC for preventing redeposition of colorant during stonewashing, and for
CC processing of textiles where cellulose breakdown is required. The new
CC truncated enzymes show reduced redeposition of dye compared to using
CC non-truncated cellulase compositions.

SQ sequence 6415 BP; 2067 A; 1082 C; 1689 G; 1541 T; 36 other;

| | | | | |
|---------------------------|-------|--------------------|------------|--------------|
| Query Match | 2.7% | Score 98.8; | DB 20; | Length 6415; |
| Best Local Similarity | 52.3% | Pred. No. 1.8e-08; | | |
| Matches 320; Conservative | 0; | Mismatches 277; | Indels 15; | Gaps 4; |

| Accession | Sequence | Position |
|-----------|---|----------|
| OY 1655 | cgcgcgtgcgcgtctccagagctcagaccgcgtctctcgtgcgcgagcccgagctccgagccgga | 1714 |
| Db 2557 | caccaacatcaacacacacacacacaccccgagccccaacacacacagtgagacgaacgcgga | 2618 |
| OY 1715 | gtagctgcgcgtctgcgcgtctccgagctcgctgcgcggtgctcg---gtgggctgaaggtgc | 1711 |
| Db 2617 | cgcgcgactccacccgcgacacgcagcggtgtccactgttcagcggaagtgtgtgaaggtac | 2676 |
| OY 1772 | agtcacgaacaaatgatctgcgcgcgcgtgataaccagatcaaacccggtctccacgttgc | 1833 |
| Db 2677 | tgctacgaacaaatgagacacagctgcgagccaggtctctctaagcgccgtgttttaagataag | 2736 |

[illegible]

| XX | RESULT | 13 |
|----|---------------|--|
| XX | AAD26526 | |
| XX | ID | AAD26526 standard; DNA; 6416 BP. |
| XX | AC | AAD26526; |
| XX | DT | 26-MAR-2002 (first entry) |
| XX | DE | Active cellulase protein, cele gene. |
| XX | OS | Unidentified. |
| XX | Key | Location/Qualifiers |
| XX | CDS | 634..5889 |
| XX | FT | /*tag= a |
| XX | FT | /product= "Cele protein" |
| XX | FT | 748..2076 |
| XX | FT | /*tag= b |
| XX | FT | /product= "DNA encoding E1 protein" |
| XX | FT | 748..2538 |
| XX | FT | /*tag= c |
| XX | FT | /product= "DNA encoding E1/2 protein" |
| XX | US6294366-B1. | |
| XX | XX | |
| XX | PD | 25-SEP-2001. |
| XX | PF | 19-AUG-1998; 98US-0136574. |
| XX | PR | 19-SEP-1997; 97US-0932571. |
| XX | PA | (CLRN) CLARIANT FINANCE BVI LTD. |
| XX | PI | Farrington GK, Anderson P, Bergquist P, Daniels R, Gibbs MD; |
| XX | PI | Morgan H, Williams DP; |
| XX | XX | |

DR WP1: 2002-081780/11.
DR P-PSDB: AAE16324.
XX
XX New cellulase active protein, useful in textile processing or
PT commercial detergents, e.g. for improving the feel or appearance of
PT cotton-containing fabrics, is stable under conditions of alkaline pH
PT and elevated temperatures -
XX
XX
XX
XX Disclosure: Column 37-44; 61pp; English.
PS
XX
XX The present invention relates to a cellulase active protein, which is
CC substantially free of proteinases of native thermophilic and
CC alkaliphilic origin, where the cellulase active protein consists of the
CC CcB5 amino acid sequence. The cellulase active protein is useful for
CC treating cellulosic materials including cotton-containing fabrics, as
CC detergent additives. The cellulase active protein is also useful for
CC improving the feel and/or appearance of cotton-containing fabrics, for
CC removing surface fibers from cotton-containing knits or for imparting
CC stone-washed appearance to cotton-containing denims. The present proteins
CC are stable under condition of alkaline pH and elevated temperatures,
CC thus suitable for textile processing and in commercial detergents.
CC The present sequence is cell gene.
XX
XX Sequence 6416 BP; 2068 A; 1082 C; 1689 G; 1541 T; 36 other;

[illegible]

| Query Match | 2.6% | Score 97.6 | DB 19 | Length 1621 |
|-----------------------|---|-------------------|----------|-------------|
| Best Local Similarity | 53.9% | Pred. No. 2.8e-08 | | |
| Matches 249 | Conservative 0 | Mismatches 204 | Indels 9 | Gaps 22 |
| 3226 | caaggcctgagcagcaactcgcttcacgcttcctgacgagcgagcagcgtccagccgac | 3285 | | |
| 1128 | ccgggtgagcgccagcagcgatcgatcgcgcgccagctcgctccgagcgagcctac | 1187 | | |
| 3286 | ccgagcccgagccgagccgagccgagccgagcccgagccgagccgcttcgagccgagc | 3345 | | |
| 1188 | gagatggcgatcgccgcgagcgagccacccaacccgagccacccgagccgctccac | 1247 | | |

[illegible]

| | |
|-------------|---|
| RESULT | 15 |
| AA\$46648/c | |
| ID | AA\$46648 standard; DNA; 6510 BP. |
| AC | |
| XX | AA\$46648; |
| XX | |
| DT | 18-DEC-2001 (first entry) |
| XX | |
| DE | Tumour suppressor gene derived chemically modified sequence #370. |
| XX | |
| KW | Human; tumour suppressor gene; oncogene; antitumour; cytostatic; |
| KW | cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP; |
| KW | cytosine methylation; ds. |
| XX | |
| OS | Homo sapiens. |
| XX | |
| PN | WO200168912-A2. |
| XX | |
| PD | 20-SEP-2001. |
| XX | |
| PF | 15-MAR-2001; 2001WO-EP02955. |
| XX | |
| PR | 15-MAR-2000; 2000DE-1013847. |
| PR | 06-APR-2000; 2000DE-1019058. |
| PR | 07-APR-2000; 2000DE-1019173. |
| PR | 30-JUN-2000; 2000DE-1032529. |
| PR | 01-SEP-2000; 2000DE-1043826. |
| XX | |
| PA | (EPIG-) EPIGENOMICS AG. |
| XX | |
| PI | Olek A, Piepenbrock C, Berlin K; |
| XX | |
| DR | WPI; 2001-602752/68. |
| XX | |
| PT | Fragments of chemically modified genes associated with tumour suppressor |
| PT | genes and oncogenes, useful in designing primers and probes for |
| PT | analysing diseases associated with cytosine methylation state e.g. |
| XX | cancer |
| XX | |
| PS | Claim 1; SEQ ID No 370; 27pp; English. |
| XX | |
| CC | The invention relates to a nucleic acid comprising a sequence of 18 |
| CC | bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with |
| CC | disulphite, of genes associated with tumour suppression and |
| CC | oncogenes having a sequence taken from 536 (actually 533 since |
| CC | numbers 408, 436 and 500 are missing from the sequence listing) sequences |
| CC | (58) and sequences complementary to (58). The nucleic acid may be a |
| CC | peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may |
| CC | |

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 31, 2002, 08:58:18 ; Search time 2693.99 Seconds
(without alignments)
18471.959 Million cell updates/sec

Title: US-09-917-384-2

Perfect score: 3687
Sequence: 1 atgagcgcgaaccacacacatc.....tcagctgcacagccagctga 3687

Scoring table: IDENTITY_NUC
Gapop 10-0, Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estbta:*
2: em_estbhum:*
3: em_estin:*
4: em_estmu:*
5: em_estrov:*
6: em_estrpl:*
7: em_estro:*
8: em_hlc:*
9: gb_estl:*
10: gb_est2:*
11: gb_hlc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1 | 99.6 | 2.7 | 1056 | 12 | CNS02URL |
| 2 | 99.4 | 2.7 | 966 | 12 | CNS02SPN |
| 3 | 97.4 | 2.6 | 906 | 12 | CNS01S6G |
| 4 | 96.8 | 2.6 | 986 | 12 | CNS01YTT |
| 5 | 93.2 | 2.5 | 971 | 12 | CNS044RS |
| 6 | 91.8 | 2.5 | 475 | 12 | CNS04S6N |
| 7 | 90.8 | 2.5 | 723 | 12 | CNS01UT4 |
| 8 | 90.8 | 2.5 | 1061 | 12 | CNS03Y86 |
| 9 | 89.2 | 2.4 | 899 | 12 | CNS02F22 |
| 10 | 86 | 2.3 | 880 | 12 | CNS01VOE |
| 11 | 85.4 | 2.3 | 803 | 12 | CNS026N7 |
| 12 | 85.2 | 2.3 | 613 | 12 | CNS02D1N |
| 13 | 84.2 | 2.3 | 949 | 12 | CNS031R8 |
| 14 | 84 | 2.3 | 559 | 12 | BH306809 |
| 15 | 83 | 2.3 | 744 | 9 | BH611425 |
| 16 | 82.6 | 2.2 | 460 | 12 | CNS0360U |
| 17 | 82.6 | 2.2 | 873 | 9 | BE217030 |

| | | | | | | |
|------|------|-----|------|----|----------|---------------------|
| C 18 | 82 | 2.2 | 614 | 12 | CNS01MDO | AL170277 Tetradon |
| C 19 | 81.6 | 2.2 | 890 | 12 | CNS035KB | AL228835 Tetradon |
| C 20 | 81.6 | 2.2 | 1101 | 12 | CNS04LE7 | AL296008 Tetradon |
| C 21 | 81.2 | 2.2 | 955 | 12 | CNS02Y44 | AL219397 Tetradon |
| C 22 | 79.2 | 2.1 | 965 | 12 | CNS01ZDK | AL174161 Tetradon |
| C 23 | 79 | 2.1 | 616 | 12 | CNS03VYH | AL262934 Tetradon |
| C 24 | 78.4 | 2.1 | 403 | 12 | CNS04F5S | AL287929 Tetradon |
| C 25 | 77.6 | 2.1 | 441 | 9 | BE125089 | BE125089 DGL_16_C1 |
| C 26 | 77.4 | 2.1 | 611 | 12 | BH379958 | BH379958 AG-ND-102 |
| C 27 | 77.2 | 2.1 | 478 | 10 | BE496065 | BE496065 WHE1261_G |
| C 28 | 77.2 | 2.1 | 523 | 10 | BE145280 | BE145280 WHE1834_F |
| C 29 | 76.8 | 2.1 | 822 | 12 | AG154144 | AG154144 Pan tlog1 |
| C 30 | 76.6 | 2.1 | 760 | 12 | CNS01NAQ | AL151987 Anopheles |
| C 31 | 76.4 | 2.1 | 401 | 10 | BE332633 | BE332633 us25402_Y |
| C 32 | 76.2 | 2.1 | 745 | 12 | CNS01N20 | AL151673 Anopheles |
| C 33 | 75.8 | 2.1 | 640 | 10 | BE427339 | BE427339 PSR6267_I |
| C 34 | 75.8 | 2.1 | 925 | 12 | CNS0091P | AL053013 Drosophila |
| C 35 | 75.6 | 2.1 | 461 | 9 | AW922188 | AW922188 DGL_1346_F |
| C 36 | 75.6 | 2.1 | 935 | 12 | CNS006XK | AL066051 Drosophila |
| C 37 | 74.2 | 2.0 | 665 | 10 | B1722807 | B1722807 1031064C0 |
| C 38 | 74 | 2.0 | 823 | 9 | AM448279 | AM448279 BRY 1782 |
| C 39 | 74 | 2.0 | 829 | 12 | CNS035GN | AL228704 Tetradon |
| C 40 | 73.4 | 2.0 | 625 | 10 | BG103499 | BG103499 RH122_37_ |
| C 41 | 72.8 | 2.0 | 710 | 12 | CNS022S1 | AL178570 Tetradon |
| C 42 | 72.6 | 2.0 | 663 | 10 | BF483682 | BF483682 WHE2336_F |
| C 43 | 72.4 | 2.0 | 554 | 12 | AZ937136 | AZ937136 2M0195012 |
| C 44 | 72.2 | 2.0 | 654 | 10 | BG367845 | BG367845 HVSME1001 |
| C 45 | 72.2 | 2.0 | 1101 | 12 | CNS017SY | AL108460 Drosophila |

ALIGNMENTS

| | | | | | |
|------------|---|------------|-----|--------|-----------------|
| RESULT 1 | CNS02URL | 1056 bp | DNA | linear | GSS 15-MAY-2000 |
| LOCUS | Tetradon nigroviridis genome survey sequence T7 end of clone | | | | |
| DEFINITION | 167F12 of library G from Tetradon nigroviridis, genomic survey sequence. | | | | |
| ACCESSION | AL214842 | GI:7873661 | | | |
| VERSION | GSS: genome survey sequence. | | | | |
| KEYWORDS | Tetradon nigroviridis. | | | | |
| SOURCE | Tetradon nigroviridis | | | | |
| ORGANISM | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Perciformes; Tetraodontiformes; Tetraodontidae; Tetradon. | | | | |
| REFERENCE | 1 (bases 1 to 1056) | | | | |
| AUTHORS | Roest-Crolius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J. | | | | |
| TITLE | Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetradon nigroviridis | | | | |
| JOURNAL | Unpublished | | | | |
| REFERENCE | 2 (bases 1 to 1056) | | | | |
| AUTHORS | Roest-Crolius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J. | | | | |
| TITLE | Human gene number estimate provided by genome wide analysis using Tetradon nigroviridis DNA sequence | | | | |
| JOURNAL | Unpublished | | | | |
| REFERENCE | 3 (bases 1 to 1056) | | | | |
| AUTHORS | Genoscope. | | | | |
| TITLE | Direct Submission | | | | |
| JOURNAL | Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases | | | | |
| COMMENT | This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetradon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetradon . | | | | |
| FEATURES | source | | | | |
| | 1..1056 | | | | |
| | /organism="Tetradon nigroviridis" | | | | |

| | | | | | | |
|--|--|------|--|--|--|--|
| TITLE | Bernot,A., Filzmes,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissbach,J. Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence Unpublished 3 (bases 1 to 906) | | | | | |
| JOURNAL REFERENCE | Genoscope. | | | | | |
| AUTHORS | Direct Submission | | | | | |
| JOURNAL TITLE | Submitted (12-APR-2000) to the EMBL/GenBank/DDBB databases | | | | | |
| JOURNAL COMMENT | This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/tetraodon.location/Qualifiers | | | | | |
| FEATURES | source | | | | | |
| BASE COUNT | 160 a 302 c 238 g 188 t 18 others | | | | | |
| ORIGIN | PUC-Or1" | | | | | |
| Query Match | 2.6%; Score 97.4; DB 12; Length 906; | | | | | |
| Best Local Similarity | 57.0%; Pred. No. 0.00062; | | | | | |
| Matches 176; Conservative 1; Mismatches 132; Indels 0; Gaps 0; | | | | | | |
| Oy 1465 | cgcagtcgctcggtcgcgagtcgcgaagcccgagtcgcgagcagtcgccatcg | 1524 | | | | |
| Db 899 | CCGAGCGCTGARTCTGAGGCTCAGGCCGGGGGTGAATCCGAGCCGGGCTGAGGCTGAGG | 840 | | | | |
| Oy 1525 | ccgtgcgcgtctcgcagagtcgcagccgcgtctcgttcgcgcgacccgcagtcgcgaccgagt | 1584 | | | | |
| Db 839 | CCGGGCGTCACTCCGAGCCCGGGGCTGAGGCCGGGGCTGAGGCTGAGTCCTGAGCTGACT | 780 | | | | |
| Oy 1585 | agctgcgcgtcgcgcgtctcgcagtcgcgaagccgcgtctcgttcgcgcgacccgcagtcgagc | 1644 | | | | |
| Db 779 | CCGAGGCCGGGACCGAGGCCGGGACCGAGGCTGAGGCTGAGGCCGGGGCTGAGGCCGGGGCGGAGG | 720 | | | | |
| Oy 1645 | cgcagtagtctgcgcgtcgcgtctcgcagtcgcgaagccgcgtctcgttcgcgcgacccgagt | 1704 | | | | |
| Db 719 | CCGGGCGTGAAGCCCGGGCTGAGGCTGAGTCCTGAGGCCGGGGCTGAGGCTGAGGCTGAGG | 660 | | | | |
| Oy 1705 | ccgagcccgagtagctgcgcgtcgcgcgtctcgcagtcgcgttcgcgcgttgctggtagctg | 1764 | | | | |
| Db 659 | CCGAGCGTGAGGCTGAGTCGAGGCCGGGGGTGAGGTTGGGGCTGAGGCCGGGGCGCGGGG | 600 | | | | |
| Oy 1765 | aagtgtag 1773 | | | | | |
| Db 599 | CCGAGGCTG 591 | | | | | |
| RESULT 4 | | | | | | |
| CNSO1YTT/c | | | | | | |
| LOCUS | 986 bp DNA linear GSS 12-MAY-2000 | | | | | |
| DEFINITION | Tetraodon nigriviridis genome survey sequence T7 end of clone 219103 of library G from Tetraodon nigriviridis, genomic survey sequence. | | | | | |
| ACCESSION | AL173450 | | | | | |
| VERSION | AL173450.1 GI:7811507 | | | | | |
| KEYWORDS | GSS; genome survey sequence. | | | | | |
| SOURCE | Tetraodon nigriviridis. | | | | | |
| ORGANISM | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon. | | | | | |
| REFERENCE | I (bases 1 to 986) Roest-Crolius,H., Jallion,O., Dasilva,C., Filzmes,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissbach,J. | | | | | |

[illegible]

[illegible][illegible]

[illegible][illegible]

[illegible][illegible]

| | | | |
|--|------|--|------|
| Oy | 1544 | cgaagccgtcttcacgctcgagcggagatccgagaccgagtagctcgccgcgtccgcttc | 1603 |
| Dd | 712 | TGAGGCTGAGGCTGAAGCTGAGGCCGGGGGCTGAGTGTGATGCCGGGGCGAGTCTGAATC | 653 |
| Oy | 1604 | cgaagctcgagccgcttctcgttcgcgcgagaccgagatccgagaccgagtagctcgccgttcg | 1663 |
| Dd | 652 | CGGGGCTGAGGCCGAGAGCTGAGCCCGGGGCTGAGGCTTGAGGCTGAGGCTGAGCCGGGGC | 593 |
| Oy | 1664 | cgtctcgagatcgagaccgctctcgttcgcgcgagaccgagatccgagaccgagtagctcg | 1723 |
| Dd | 592 | TGAGTGTGATGTCGGGGGCTGAGTGTGACTCCGGGGCTGAGGCTGAGGCTGAGGC | 533 |
| Oy | 1724 | cgctgcgcgttcgcgagctcgttcgcgcgagaccgagatccgagaccgagtagctcg | 1757 |
| Dd | 532 | CGAGGCTGAGGCGCAGGCTGAGGCCGGGGCTGAG | 499 |
| RESULT 11 | | | |
| CNS026N7 803 bp DNA linear GSS 12-MAY-2000 | | | |
| LOCUS CNS026N7 | | | |
| DEFINITION Tetradodon nigraviridis genome survey sequence PUC-Orl end of clone | | | |
| 241C10 of library G from Tetradodon nigraviridis, genomic survey | | | |
| sequence. | | | |
| ACCESSION ALI83580 | | | |
| VERSION ALI83580.1 GI:7821684 | | | |
| KEYWORDS GSI; genome survey sequence. | | | |
| SOURCE Tetradodon nigraviridis. | | | |
| ORGANISM Tetradodon nigraviridis | | | |
| Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | |
| Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; | | | |
| Acanthomorpha; Acanthopterygii; Percomorphia; Tetraodontiformes; | | | |
| Tetraodontidae; Tetradodon. | | | |
| REFERENCE 1 (bases 1 to 803) | | | |
| AUTHORS Roest-Crollius,H., Jalllon,O., Dasilva,C., Fizames,C., Fisher,C., | | | |
| Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and | | | |
| Weissenbach,J. | | | |
| TITLE Characterization and repeat analysis of the compact genome of the | | | |
| JOURNAL freshwater pufferfish Tetradodon nigraviridis | | | |
| REFERENCE Unpublished | | | |
| AUTHORS 2 (bases 1 to 803) | | | |
| Roest-Crollius,H., Jalllon,O., Dasilva,C., Bouneau,L., Fisher,C., | | | |
| Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., | | | |
| Saurin,W. and Weissenbach,J. | | | |
| TITLE Human gene number estimate provided by genome wide analysis using | | | |
| JOURNAL Tetradodon nigraviridis DNA sequence | | | |
| REFERENCE Unpublished | | | |
| AUTHORS 3 (bases 1 to 803) | | | |
| Genoscope. | | | |
| DIRECT SUBMISSION Direct Submission | | | |
| JOURNAL Submitted (12-APR-2000) to the EMBL/Genbank/DBJ databases | | | |
| COMMENT This sequence is a single read and was generated as part of a large | | | |
| scale clone-and sequencing project of the Tetradodon nigraviridis | | | |
| genome. For more information, please take a look at | | | |
| http://www.genoscope.cns.fr/Tetraodon. | | | |
| FEATURES | | | |
| source | | | |
| 1..803 | | | |
| /organism="Tetradodon nigraviridis" | | | |
| /db_xref="taxon:99883" | | | |
| /clone="241C10" | | | |
| /clone_1lb="G- | | | |
| /note="Genoscope sequence ID : CGAG241B05SP1-end : | | | |
| PUC-Orl" | | | |
| BASE COUNT 95 a 218 c 223 g 192 t 75 others | | | |
| ORIGIN | | | |
| Query Match 2.3%; Score 85.4; DB 12; Length 803; | | | |
| Best Local Similarity 52.1%; Pred.No.0.024; | | | |
| Matches 139; Conservative 18; Mismatches 110; Indels 0; Gaps 0; | | | |
| Oy | 1449 | ccccgcgcctccgccgcgcgagatcgcttggttcgccgagatccgagaccgagatccgagccc | 1508 |
| Dd | 235 | CACGGGGCTCCAGCGCTGACGAGCGTCCGCAAGDAGGCCCGCTGAAVCTGNSCCGACCCT | 294 |

[illegible]

| Accession | Sequence | Position |
|-----------|--|----------|
| QY 1512 | cagctcgcacatcgccgtgcgcgtccagctccagatccagccggtctccgtccgcagccag | 1571 |
| Db 237 | GCTGAGGCGCGGGGCTGAGAGCTGAGTCCGGGGGTGAGGCTGAGAGCTGAGGCTGAGGCCGGG | 296 |
| QY 1572 | tcgcagaccgagtagctgcgcgtgcgcgtctccagatccagaccggtctccgtgcgcag | 1631 |
| Db 297 | GCTGAGGCTGAGTCCGGGGCTGAGGCTCGAGGCTGAGGCTGAGGCCGCGGGGCTGAGGCTGAG | 356 |
| QY 1632 | cccgagctccgagcccgagtagctgcgcgtgcgcgtccagagctcgagcccgctccgtc | 1691 |
| Db 357 | TCCGGGGGCGGGGCGCTGAGAGCTGAGGCTGAGGCTGAGGCTGAGGCCGGGGCTGAGGCTGAG | 416 |
| QY 1692 | gcccagaccgagctccgagcccgagtagctgcgcgtgcgcgtctccagagctgcgtccggt | 1751 |
| Db 417 | GCTGAGGCGCGGGGCTGAGAGCGGAGGCTGAGGCCGGGGGCTGAGAGCTGAGGCTGAGGCTGAG | 476 |
| QY 1752 | gtccgag 1757 | |
| Db 477 | GCCGGG 482 | |

| | |
|------------|---|
| RESULT | 13 |
| CNS031R8 | |
| LOCUS | 949 bp DNA linear GSS 15-MAY-2000 |
| DEFINITION | Tetradon nigroviridis genome survey sequence T7 end of clone 205022 of library G from Tetradon nigroviridis, genomic survey sequence. |
| ACCESSION | AF223901 |
| VERSION | AF223901.1 GI:7882739 |
| KEYWORDS | GSS; genome survey sequence. |
| SOURCE | Tetradon nigroviridis. |
| ORGANISM | Tetradon nigroviridis |

REFERENCE
AUTHORS
1 (bases 1 to 949)
Roest-Crollius, H., Jallion, O., Dasilva, C., Fizames, C., Fisher, C.,

TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish *Tetraodon nigroviridis*

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 949)

AUTHORS Roest-Criollus, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C.

| | |
|-----------|---|
| TITLE | Saurin, W. and Weissensech, J. Human gene number estimate provided by genome wide analysis using Tetradon nigroviridis DNA sequence |
| JOURNAL | Unpublished |
| REFERENCE | 3 (bases 1 to 949) |

JOURNAL COMMENT
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
This sequence is a single read and was generated as part of a larger

genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

| FEATURES | SOURCE | Location/Qualifiers |
|------------|--|----------------------------|
| | 1. .949 | |
| | /organism="Telraodon nigriviridis" | |
| | /db_xref="taxon:39883" | |
| | /clone="205022" | |
| | /clone_11b="G" | |
| | /note="Genoscope sequence ID : C06G205B11bP1-end : T7" | |
| BASE COUNT | 205 a | 202 c 390 g 144 t 8 others |
| ORIGIN | | |

| | | | | |
|---------------------------|-------|-----------------|-----------|-------------|
| Query Match | 2.3% | Score 84.2 | DB 12, | Length 949; |
| Best Local Similarity | 49.0% | Pred. No. 0.034 | | |
| Matches 221; Conservative | 1; | Mismatches 229; | Indels 0; | Gaps 0; |

[illegible]

| | | | | | |
|------------|---|-------------|-----------|-------------------|-----------------|
| RESULT | 14 | | | | |
| LOCUS | BH306809 | | | | |
| DEFINITION | BH306809 | 559 bp | DNA | linear | GSS 30-NOV-2001 |
| ACCESSION | CH230-100P7.TV | CHOR1-230 | Segment 1 | Rattus norvegicus | genomic clone |
| VERSION | CH230-100P7, | | | | |
| KEYWORDS | BH306809.1 | GI:17219217 | | | |
| SOURCE | GSS. | | | | |
| ORGANISM | Norway rat. | | | | |
| | Rattus norvegicus | | | | |
| | Euryarchaea; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | | |
| | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; | | | | |
| | Rattus. | | | | |
| REFERENCE | 1 (bases 1 to 559) | | | | |
| AUTHORS | Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de Jong,P., and Fraser,C.M. | | | | |
| TITLE | Rat BAC End Sequences from Library CHOR1-230 ECORI segment | | | | |
| JOURNAL | Unpublished (1999) | | | | |
| COMMENT | Other_GSSs: CH230-100P7.TJ | | | | |

Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhac@tigr.org
Clones are derived from the rat BAC library CHORI-230
(<http://www.chori.org/Dacpac/Fat230.htm>). For BAC library
availability, please contact Pletier de Jong (pdejong@email.cho.org)

http://www.chori.org/bacpac/or_e1ing_information.html. BAC end
 page: http://www.f1gr.org/ribd/Bac_ends/rat/bac_end_intro.html
 Plate: 100 row: p column: 7
 Seq primer: T7
 Class: BAC ends.

```

FEATURES      Location/Qualifiers
source        1. .559
              /organism="Rattus norvegicus"

```

[illegible][illegible]

Tue Sep 3 13:41:59 2002

us-09-917-384-2.rst

Page 10

Db 416 CCG 418

Search completed: August 31, 2002, 10:40:44
job time: 6146 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 31, 2002, 09:04:38 ; Search time 77.71 Seconds
(without alignments)
11654.237 Million cell updates/sec

Title: US-09-917-384-2
Perfect score: 3687
Sequence: 1 atgagagcgaaccacacatc.....tcagctgcacagcagctga 3687

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Issued_Patents_NA:*

1: /cgn2_6/ptodata/1/lna/5A_COMB.seq:*\n2: /cgn2_6/ptodata/1/lna/5B_COMB.seq:*\n3: /cgn2_6/ptodata/1/lna/6A_COMB.seq:*\n4: /cgn2_6/ptodata/1/lna/6B_COMB.seq:*\n5: /cgn2_6/ptodata/1/lna/PCTUS_COMB.seq:*\n6: /cgn2_6/ptodata/1/lna/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|---------|---------------------|--------------------|
| 1 | 427 | 11.6 | 3004 | 1 US-08-276-213-6 | Sequence 6, Appl1 |
| 2 | 324 | 8.8 | 3404 | 1 US-08-265-429A-1 | Sequence 1, Appl1 |
| 3 | 324 | 8.8 | 3404 | 1 PCT-US95-09069-1 | Sequence 1, Appl1 |
| 4 | 314.6 | 8.5 | 1269 | 1 US-08-265-429A-4 | Sequence 4, Appl1 |
| 5 | 314.6 | 8.5 | 1269 | 5 PCT-US95-09069-4 | Sequence 4, Appl1 |
| 6 | 134.4 | 3.6 | 11707 | 4 US-09-136-574A-1 | Sequence 1, Appl1 |
| 7 | 116.4 | 3.2 | 2977 | 2 US-07-862-588B-1 | Sequence 1, Appl1 |
| 8 | 98.8 | 2.7 | 6416 | 4 US-09-136-574A-2 | Sequence 2, Appl1 |
| 9 | 95.2 | 2.6 | 1116 | 4 US-09-104-308-2 | Sequence 2, Appl1 |
| 10 | 95.2 | 2.6 | 1116 | 4 US-09-321-981-2 | Sequence 2, Appl1 |
| 11 | 94.8 | 2.4 | 390 | 4 US-09-197-649-7 | Sequence 7, Appl1 |
| 12 | 87.8 | 2.4 | 390 | 4 US-09-197-649-7 | Sequence 7, Appl1 |
| 13 | 87.8 | 2.4 | 4403765 | 4 US-09-103-840A-2 | Sequence 2, Appl1 |
| 14 | 81.6 | 2.2 | 4403765 | 4 US-09-103-840A-2 | Sequence 2, Appl1 |
| 15 | 81.2 | 2.2 | 2029 | 4 US-09-136-574A-6 | Sequence 4, Appl1 |
| 16 | 80 | 2.2 | 4411529 | 4 US-09-103-981A-1 | Sequence 1, Appl1 |
| 17 | 80 | 2.2 | 1716 | 4 US-09-321-981-4 | Sequence 4, Appl1 |
| 18 | 77.4 | 2.1 | 2712 | 3 US-09-025-691-4 | Sequence 4, Appl1 |
| 19 | 77.4 | 2.1 | 985 | 4 US-09-056-556-182 | Sequence 182, App |
| 20 | 77 | 2.1 | 2293 | 1 US-08-604-913B-12 | Sequence 12, Appl1 |
| 21 | 74.2 | 2.0 | 1931 | 2 US-09-130-114-2 | Sequence 1, Appl1 |
| 22 | 70.4 | 1.9 | 4411529 | 4 US-09-103-840A-1 | Sequence 1, Appl1 |
| 23 | 69.4 | 1.9 | 1280 | 4 US-09-060-756-4 | Sequence 4, Appl1 |
| 24 | 68.2 | 1.8 | 535 | 4 US-09-056-556-171 | Sequence 171, App |
| 25 | 67.4 | 1.8 | 897 | 4 US-09-434-288-6 | Sequence 6, Appl1 |
| 26 | 67.4 | 1.8 | 2668 | 2 US-08-461-775-11 | Sequence 11, Appl1 |
| 27 | 67.4 | 1.8 | 2668 | 3 US-09-031-606-11 | Sequence 11, Appl1 |

| | | | | | |
|------|------|-----|-------|---------------------|--------------------|
| C 28 | 66.8 | 1.8 | 494 | 4 US-09-056-556-176 | Sequence 176, App |
| 29 | 65.8 | 1.8 | 1248 | 4 US-09-105-537-7 | Sequence 7, Appl1 |
| C 30 | 65.8 | 1.8 | 13613 | 4 US-09-105-537-3 | Sequence 3, Appl1 |
| 31 | 65.4 | 1.8 | 2219 | 3 US-08-510-646B-17 | Sequence 17, Appl1 |
| C 32 | 65.4 | 1.8 | 5970 | 3 US-09-320-878-21 | Sequence 21, Appl1 |
| 33 | 65.4 | 1.8 | 28958 | 1 US-08-238-261B-6 | Sequence 6, Appl1 |
| 34 | 65.4 | 1.8 | 28958 | 1 US-08-456-837-6 | Sequence 6, Appl1 |
| 35 | 65.4 | 1.8 | 28958 | 1 US-08-457-342-6 | Sequence 6, Appl1 |
| 36 | 65.4 | 1.8 | 28958 | 1 US-08-457-646A-6 | Sequence 6, Appl1 |
| 37 | 65.4 | 1.8 | 28958 | 1 US-08-458-076A-6 | Sequence 6, Appl1 |
| 38 | 65.4 | 1.8 | 28958 | 1 US-08-764-233A-4 | Sequence 4, Appl1 |
| 39 | 65.4 | 1.8 | 28958 | 1 US-08-457-335A-6 | Sequence 6, Appl1 |
| 40 | 65.4 | 1.8 | 28958 | 1 US-08-729-214-6 | Sequence 6, Appl1 |
| 41 | 65.4 | 1.8 | 28958 | 3 US-09-028-934-6 | Sequence 6, Appl1 |
| 42 | 65.4 | 1.8 | 49377 | 1 US-08-764-233A-1 | Sequence 1, Appl1 |
| 43 | 65.2 | 1.8 | 43280 | 2 US-08-804-227C-1 | Sequence 1, Appl1 |
| 44 | 64.8 | 1.8 | 15872 | 4 US-09-105-537-1 | Sequence 1, Appl1 |
| 45 | 64.8 | 1.8 | 44377 | 2 US-08-804-227C-7 | Sequence 7, Appl1 |

ALIGNMENTS

RESULT 1
US-08-276-213-6
Sequence 6, Application US/08276213
Patent No. 553655
GENERAL INFORMATION:
APPLICANT: Thomas, Steven
APPLICANT: Laymon, Robert
TITLE OF INVENTION: GENE ENCODING FOR THE E1 ENDOGLUCANASE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: National Renewable Energy Laboratory
STREET: 1617 Cole Boulevard
CITY: Golden
STATE: CO
COUNTRY: USA
ZIP: 80401-3393
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/276.213
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: O'Connor, Edna
REGISTRATION NUMBER: 29,252
REFERENCE/DOCKET NUMBER: NREL IR# 94-08
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303)231-1000
TELEFAX: (303)231-1098
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 3004 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-276-213-6
Query Match 11.6%; Score 427; DB 1; Length 3004;
Best Local Similarity 99.8%; Pred. No. 1.6e-73;
Matches 438; Conservative 0; Indels 1; Gaps 1;

```

OY 1 atggagcgaaccacaatccgagcgaagtcaggtaccagagaggaacagacgaatg 60
   |||
Db 2567 ATGGAGCGAACCACCAATCCGAGCGAAGTGCAGGTACCGAGAGAAAGACAGCAATG 2626
OY 61 cccgcacatcacaacgagctgagcgagcgctcctcgcggcgagctgagcatcgagcc 120
   |||
Db 2627 CCCGCCATCTCAAAAGCGCTGCGAGCGGCGTCCCTCGCCGGGGCGGTGAGCATCGCAGCC 2686
OY 121 tccatcgctgcgctgagctgagcatcctcgtccatcgcgcgagcagcagctgacaatccc 180
   |||
Db 2687 TCCATCGTCCCTCGCGAGTGCAGTCCTGCCATGCGCGAGCGACATCGCAATGCC 2746
OY 181 tatcgaggagcagcctcttcgtcaaccgtactggcgcaagaagtacagaagcagaag 240
   |||
Db 2747 TATGCGGAGGACCACTTCTTCTGTCACACCGTACTGGGCGCCAAAGAACTAGAGCGAA -CG 2805
OY 241 gcgaacagacaatgacatcctcgcagcgaaatgagcgcttcacataatcgagc 300
   |||
Db 2806 GCGAACCAACCAATGCCACTCTCGCAGCGAAATCGCGGTTCACATATTCGAGC 2865
OY 301 gccgctctgtagcgagcagctcgtcgatcaaacgagcgtaaacggcgagccgcttgaag 360
   |||
Db 2866 GCGGCTGATGAGCGACGTCGCTGCGATCAACGGCGTCAACGGCGAGCCCGCTTGAGC 2925
OY 361 acatatctgagcgagcctctccagcagcagcagcagcagcagcagcagcagcagcagc 420
   |||
Db 2926 ACATATCTGAGAGCGCCGCTCTCCAGCAGCAGGAAACCAACCCCTGAAGTCAATGAGATT 2985
OY 421 gtcattacgactgcgcg 439
   |||
Db 2986 GTCATCTACGATCTGCCG 3004

```

RESULT 2

```

; Sequence 1, Application US/08265429A
; Patent No. 5677151
; GENERAL INFORMATION:
; APPLICANT: Wilson, David B.
; APPLICANT: Walker, Larry P.
; TITLE OF INVENTION: Thermostable Cellulase From A Thermomonospora Gene
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hodgson, Russ, Andrews, Woods & Goodyear
; STREET: 1800 One Mt Plaza
; CITY: Bufileo
; STATE: New York
; COUNTRY: United States
; ZIP: 14203-2391
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
; COMPUTER: IBM compatible
; OPERATING SYSTEM: MS-DOS/ Microsoft Windows 3.1
; SOFTWARE: Wordperfect for Windows 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/265,429A
; FILING DATE: 24/06/94
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Nelson, M. Bud
; REGISTRATION NUMBER: 35,300
; REFERENCE/DOCKET NUMBER: 18617,0008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716)856-4000
; TELEFAX: 716-849-0349
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3404 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: double-stranded
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA

```

```

; ORIGINAL SOURCE:
; ORGANISM: Thermomonospora fusca
; STRAIN: YX36
; CELL TYPE: bacterium
; US-08-265-429A-1

```

```

Query Match 8.8%; Score 324; DB 1; Length 3404;
Best Local Similarity 56.6%; Pred. No. 8.5e-54;
Matches 775; Conservative 0; Mismatches 535; Indels 60; Gaps 7;

```

```

OY 200 tctgcaaccgctactggtgagcgaagaagtacagcggaagcggcggaaccagaccgaatgcca 259
   |||
Db 1010 TCGACAAACCCGTTCTGAGGGCGCCAACTGACTGAACCCGGTCTGTGGCCAAAGGCCG 1069
OY 260 ctctcgcaagcaaatgagcgctgttccacatatcgacgagcgctctgtagttagcagca 319
   |||
Db 1070 CCGCTGAGCGGGCGGGTCCCGCGGTGCCCAAGATGCCACCGCTGTCTGGCTGAGACGTA 1129
OY 320 tegtgcgatacaagcgctcaacgagcggaaccggcttgaagacatatctgagcgccccc 379
   |||
Db 1130 TCGGCGCATCGAGGGCAACGACGCGCCGACCCGCTTCATGGGTCTGCGGACCCACC 1189
OY 380 t-----ctccagcagcagcagcgaaccccttgaatgaattgattgattatcagc 430
   |||
Db 1190 TGGAGGAGGCGCTCCGCGCATCGGTGGCGAGACCCGCTGACATCCAGTGTATCTACA 1249
OY 431 atctcgaggagcagcagctgcgcgctcgcctcacaacgagcgaatgcgcgtacagcgag 490
   |||
Db 1250 ACCTGCCGCGCGCGACTGCGCGGCGCTGGCGCTCCCAAGGTAGACTG-----GTCGCG 1303
OY 491 caggtttgcaagacctgaagaagcagtaacatgcatcgaattgcgagatcctgagaat- 549
   |||
Db 1304 ATGAATCTGACCGCTCAAGAGCGAGTACATCATCCAGTCCGACATCATATCTGGACT 1363
OY 550 --ccgaagtaactccagcctgagcagatgtagatgaattgaacggagcagctgcgcaagc 607
   |||
Db 1364 TCGCACAATCAAGAACTCTGCGATGTCGCCATTCATCGATCGACTCCCTGCCAACCC 1423
OY 608 cggtcacacatatgagcattcaagcgtgtgcaagcgagctgcgta----- 653
   |||
Db 1424 TCGTCAACCAAGTGGCGGGAACGGCGGACCGACGAGCTGCGGCTCAATGAAGCAGAAAG 1483
OY 654 ----taagagcaagcagcagtaacgagcgtcaagaaatgcagcattcgcgaatcgcgaat 709
   |||
Db 1484 GCGGCTACGTCAACGGTGTGCGGTACGCCCTCGCAAGCTGGCGGAGATCCCAAGCTCT 1543
OY 710 acatcacatgagcgcgcgccactccgctggtgtgtggtggtggtggtggtggtggtggtggt 769
   |||
Db 1544 ACAACTACATCGAGCGCGCCACCAAGGCTGGATCGGTGGGACTCCAACTTGGCCCTT 1603
OY 770 acgtacaggaagtcagaaggtctcaagc---cgaagcactcggtgtggtggtggtggtggtggt 826
   |||
Db 1604 CGGTGACATCTCTCAAGAGGCGCCCAACGCGCTCGGCTCCACAGGAGTACGTGTCAGG 1663
OY 827 gcttgcacacaacagcggaattacagcgcttgaagaagccgttcatgagccgaccc 886
   |||
Db 1664 GCTTCATCTCCAAACAGCGCCAACTACTCGGCCCATCTGTGAGCGGTACTCGAGAGTCAAG 1723
OY 887 agcagtcgagcagcagcggtgagtcgagcgaattctacacgagtggaatccgagatcg 946
   |||
Db 1724 GCACGTTAAAGGCGCAAGCTCATCGCCCATCAAGTGGGTGACTGGAACCACTAGTGTG 1783
OY 947 acgaagcgcgactacgagctgagcttactcgcgagctgctgcgcgagcttctcaaga 1006
   |||
Db 1784 ACGAGCTCTCCTTCGTCAGAGACTCGTCAAGGCGCTGATCGCAGAGGCGCTTCCGGTCCG 1843
OY 1007 gcaatgcgcatgcatcagacacctacgcaacgagttggggtggttccgaaacgaacaa 1066
   |||
Db 1844 ACATCGGTATGCTCATCGACACTCCCGAACGCGTGGGGGTGGCCGGAACCGTCCGACCG 1903
OY 1067 gcccgagcagcgcgagcagctgtcaacacctgtcaacagctgcaagctgcaagcttgcctcgcg 1126
   |||

```

[illegible]

RESULT 3
PCT-US95-09069-1
; Sequence 1, Application PC/TUS9509065

```

:
: APPLICANT: Wilson, David B.
: APPLICANT: Walker, Larry P.
: APPLICANT: Zhang, Sheng
: TITLE OF INVENTION: Thermostable Cellulase From A Thermomonospora Gene
:
: NUMBER OF SEQUENCES: 8
:
: CORRESPONDENCE ADDRESS:
:

```

ADDRESSSEE: Hodgson, Russ, Andrews, Woods & Goodyear
STREET: 1800 One Mt Plaza
CITY: Buffalo
STATE: New York
COUNTRY: United States
ZIP: 14203-2391
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
COMPUTER: IBM compatible
OPERATING SYSTEM: MS-DOS/ Microsoft Windows 3.1
SOFTWARE: Wordperfect for Windows 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/09069

1 FILING DATE:
2 CLASSIFICATION:
3 PRIOR APPLICATION DATA:
4 APPLICATION NUMBER: U.S. Serial No. 08/265,423
5 FILING DATE: 24/06/94
6 ATTORNEY/AGENT INFORMATION:
7 NAME: Nelson, M. Bud
8 REGISTRATION NUMBER: 35,300
9 REFERENCE/DOCKET NUMBER: 18617.0008
10 TELECOMMUNICATION INFORMATION:
11 TELEPHONE: (716)856-4000
12 TELEFAX: 716-849-0349
13 INFORMATION FOR SEQ ID NO: 1:

```

;
; SEQUENCE CHARACTERISTICS:
;     LENGTH: 3404 nucleotides
;     TYPE: nucleic acid
;     STRANDEDNESS: double-stranded
;     TOPOLOGY: linear
;     MOLECULE TYPE: Genomic DNA
;

```

```

: ORIGINAL SOURCE:
: ORGANISM: Thermomonospora fusca
: STRAIN: YX36
: CELL TYPE: bacterium
: PCT-US95-09069-1

```

| | | | | |
|-----------------------|--------------|--------------------|-----------------|--------------|
| Query Match | 8.8%; | Score 324; | DB 5; | Length 3404; |
| Best Local Similarity | 56.68; | Pred. NO. 8.5e-54; | | |
| Matches 775; | Conservative | 0; | Mismatches 535; | Indels 60; |
| | | | | Gaps 7 |

| | | | |
|----|------|---|--------|
| OY | 200 | ttgtcaacccgtactctgtggcgcaagaagtacagagcgaaagcggcgagaccagacgaattgtca | 259 |
| Db | 1010 | TCGACAAACCCGTTTCGAGGGGGCCAAACGTACTGTATGAAACCCGGTCTGTGCTGGCCAAAGGCCG | 10689 |
| OY | 260 | ctctcgcagcgaaaatatgcgtcgtcgtcttccacatatctcgacggcgtctctgtatgtgacgca | 319 |
| Db | 1070 | CCGCTGAGCCGGGGCGGTTCCGGCGGTGGCCAAAGATCCACCCGCTGTCTGGCTGGCGCTGA | 11299 |
| OY | 320 | tcgctgtgatacgaacggtgtcaaaagcgcgacccggctctgacgacacatatcttgagcgcgcgcc | 379 |
| Db | 1130 | TCGGCGCGCATCGAGGGGCAAGACAGCCCGACACCCGGCTCCATGGGTGTGGCGGCACACC | 11899 |
| OY | 380 | t-----ctccagcgcaagggacacaccaccttaagtcaatttgatattgtactatcg | 430 |
| Db | 1190 | TGGAGAGAGGGCGTCCGCGCACTCCGGTGTGGGAGACCCGCTGACATCGAAGTCTGTCATCTACA | 12499 |
| OY | 431 | atctgcgggacgagactgcgcgcgcctgcctcaagcggaactgcgcgtactacgycag | 490 |
| Db | 1250 | ACCTGCGCGGCGCGACTGTGGCGCGCTGTGCCCTCCAAACGGTAGCTG-----GGTCCCG | 13039 |
| OY | 491 | caggtctgcgagactatgtaaagcgaatcatatgatcgtatgtcgatcttcctgagcaat- | 549 |
| Db | 1304 | ATGAACTCGAACCGTTACAAAGAGCGAGTACTGACCCCGATCCCGACATCATGTGGGACT | 13639 |
| OY | 550 | --cgaagctactcagacccttcgagatctgtacgatacttgacgcgagactgcgtgccaaacg | 607 |
| Db | 1364 | TCGGAGACTACGAGAAACCTGTGGATGTGTGCATATGAGATGACATCCCTGGCCCAAC | 14239 |
| OY | 608 | cgtgtaccacaaatagtacatcaacagcgtgtgcgaagcggtgtgcgtta-----653 | |
| Db | 1424 | TCGTGACCAACGTGGGGGGGAGACGGGGGACCCAGCTCTGCGCCTCATATGAAGCGAAGC | 14839 |
| OY | 654 | -----ttacggcgaaggtacgtagttacgcgtctcagaaatttgacgcacattccgaacggt | 709 |
| Db | 1484 | GCGGCTACGTCAACGGGTGTGGCTACGCCCTCCGCAAGCTGGGGCAGATCCCGAAACGCT | 15439 |
| OY | 710 | acactatcatgagacgcgcgcacactctcgctgcgttgcgtgtgcgcacaaatgacgcaagcat | 769 |
| Db | 1544 | ACAACCTACATGAGACGGCGCCCAACAGGGGTGATGTGGATACATCCAACTTCGGCCCT | 16039 |
| OY | 770 | acgtacaggaagttccagaaggtctctcaag--cgaacatcgggtgtlcaagcgatcatcgag | 826 |
| Db | 1604 | CGGTGGACATCTTTATACGAGGCGGCCCAACGCCCTCCGCTCCACCGTGGACTACGTGCACG | 16639 |
| OY | 827 | gcttcgtacacaaacagcggaattacacgcggttgaaaggacgcgttctagaccgcacc | 886 |
| Db | 1664 | GCTTCATCTCCAAACACGGCCAACTACTCGGCCACTGTGGAGGCCGTATCTTGACGATCAACG | 172239 |
| OY | 887 | agcagctcgcgcggaacgcgcgttgtagtctcgcgatcttccacagtggaatctctgacatcg | 946 |
| Db | 1724 | GCACCGTTAACGGGCACTCATCCGCCAGTCCAAAGTGGGTTGACTGGAACACGATACGTG | 17839 |
| OY | 947 | acggaagccgactaagcgggtgtgaatttactcgcgcgtcgcgcgtcgtgcttccaagca | 10066 |
| Db | 1784 | ACGAGCTCTCTCTGTGTCAAGGACTCGTCAAGGCCCTTGATGCGCCAAAGGGCTTCCGGTCCG | 18439 |
| OY | 1007 | gcatcgcgcatgtcatcgacaacttcgcaaggttgtgggtgtctcgaaacggaacaaacag | 106666 |
| Db | 1844 | ACATCGGTATGCTATGTGAACACTCCCGCAAGGGCTGGGGGTGGCCGAACGCTCCGACCG | 190303 |
| OY | 1067 | gcccgcacacgcgagacatgtlcaaacactctgtlcaacagatcgaaagtgtgaactcttggc | 11266 |

Db 1904 GACGAGCTCTCCACGACGCTCAACACCTACGTTGACGAGAGCCGTATCGACGCCGCTA 1963
Qy 1127 agcaaccgagccttgatgacaaagaaagtgacgagcctgagcagcagcagcagcagc 1186
Db 1964 TCCACCCCGGTACTGCTGCAACAGGCGGCTGCGGCGCTGCGCAAGC----- 2011
Qy 1187 gccgaacgagcttcgcaagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1246
Db 2012 GCGCCACGCTAACCCCGGCTCCCGGTGTTGACGCTTACGCTGAGTGAAGCCCGGGGTG 2071
Qy 1247 agtcgaacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1306
Db 2072 AGTCGACGAGCGGCGGAGGAGATCCGGAACGAGAGGAGGAGGAGGAGGAGGAGGAGG 2131
Qy 1307 ggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1357
Db 2132 GCGACCCGCTGCTGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2191
Qy 1358 actccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1417
Db 2192 AGCGCCCATCTCCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2251
Qy 1418 ggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1477
Db 2252 ACCCGCTCTGTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2311
Qy 1478 ctgcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1527
Db 2312 CACTCTCTGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2361

RESULT 4

US-08-265-429A-4

Sequence 4, Application US/08265429A
Patent No. 5677151
GENERAL INFORMATION:
APPLICANT: Wilson, David B.
APPLICANT: Walker, Larry P.
APPLICANT: Zhang, Sheng
TITLE OF INVENTION: Thermostable Cellulase From A Thermomonospora Gene
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hodgson, Russ, Andrews, Woods & Goodyear
STREET: 1800 One M&P Plaza
CITY: Buffalo
STATE: New York
COUNTRY: United States
ZIP: 14203-2391
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 MB storage
OPERATING SYSTEM: MS-DOS/ Microsoft Windows 3.1
SOFTWARE: Wordperfect for Windows 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/265,429A
FILING DATE: 24/06/94
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Nelson, M. Bud
REGISTRATION NUMBER: 35,300
REFERENCE/DOCKET NUMBER: 18617.0008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716)856-4000
TELEFAX: 716-849-0349
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1269 nucleotides
TYPE: nucleic acid
STRANDEDNESS: double-stranded
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
ORIGINAL SOURCE:
ORGANISM: Thermomonospora fusca

STRAIN: YX36
CELL TYPE: bacterium
US-08-265-429A-4
Query Match 8.5%; Score 314.6; DN 1; Length 1269;
Best Local Similarity 57.6%; Pred. NO. 4.7e-52;
Matches 732; Conservative 0; Mismatches 479; Indels 60; Gaps 7;
Qy 200 tgcgaacccgctactgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 259
Db 17 TCGACACCCGTTCCAGAGGCGCCCAAGCTGTACGTACACCCGCTGCTGCTGCTGCTGCTGCTG 76
Qy 260 ctctcgcagcgaataatgcgctgcttcccaatattcgaagcagcagcagcagcagcagcagcagc 319
Db 77 CCGCTGACCCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 136
Qy 320 tgcgctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 379
Db 137 TCGGCGCCATCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 196
Qy 380 t-----ctccagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 430
Db 197 TCGAGAGGCGCGTCCGCAAGTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 256
Qy 431 atctgcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 490
Db 257 ACCTCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 310
Qy 491 caggttgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 547
Db 311 ATGAACCTGACCGCTTACAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 370
Qy 548 atccgaagctactcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 607
Db 371 TCGGAGACTACGAGAACTCGGAGATCGTCCATCATGAGATGAGATGAGATGAGATGAGATGAG 430
Qy 608 cgttcaccaatattgagcattcaagcgttgtgcaagcagcagcagcagcagcagcagcagcagc 654
Db 431 TCGTCACCAACCTGGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 490
Qy 655 -----tacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 709
Db 491 GCGGCTTACGTCACAGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 550
Qy 710 acatctacatgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 769
Db 551 ACAACTTACATGAGCGCGCGCCACACAGGCTGATGCGGTGGGAGCTCAACTTCGGCCCT 610
Qy 770 acgtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 826
Db 611 CGGTGAGATCTTTCACAGAGGCGCGCCACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 670
Qy 827 gcttgatcaccaacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 886
Db 671 GCTTATCTCCACACGCGCCCAACTCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 730
Qy 887 agc 946
Db 731 GCACGCTTACAGCGCGCGCTCATCCGCAAGTGGGTTGCAAGTGGGTTGCAAGTGGGTTGCA 790
Qy 947 acgaagc 1006
Db 791 ACGACCTCTCTTCTCCAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 850
Qy 1007 gcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatg 1066
Db 851 ACATGGATGCTCATGCAACCTCCCGCAAGCGGTGGGTTGCGCAAGCGGTGGGTTGCGCAAG 910
Qy 1067 gcccgagc 1126
Db 911 GACGAGGCTCTCCACGACGACCTCAACACTACGTTGACGAGAGCGGCTATCGACGCCCGCTA 970

```

QY 1127 agcaccgagcctgtgtgtcaaccagaacggtgtgggctctggccagccgcgaggaac 1186
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 971 tccaccctccctactgtgtgtcaacacgagccgctgtggccctccgagcagccccc----- 1023
QY 1187 gcccgagcagctctccgaacgcgacacctgagcgcgfatgtctgtatcaaggccgcgggtg 1246
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1024 -----ACGGTCAACCCCGCTCCCGGTGTGTGACGCTTACGTCTGTGGTGAAGCCCGGGTg 1078
QY 1247 agtcgagcagcacaacgctgtcgagcgatcgcacactgtgcaagaagtcgagcccatgt 1306
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1079 actcgagcagccgcacacgagagatccgacagcagagggccttcgacgcgcatgt 1138
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1307 ggcagccgacgtaacacgagctgtacggtgtactgacccaac-----gggtaccga 1357
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1139 gcgacccgacactacacgagccacacgcccgcacagcgaacacccctcggtgcgctccca 1198
QY 1358 actcccgagtcgagcagcagtggttccggcgagtttgacagctgtgtcggaagcaac 1417
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1199 acgccccacatctccgcccactgtgtcttgcctcagttccgagctgtgcccacacgctt 1258
QY 1418 ggcagcaggtg 1428
    || | || |
DB 1259 acccgctctgtg 1269

```

```

RESULT 5
PCT-US95-09069-4
: Sequence 4, Application PC/TUS9509069
: GENERAL INFORMATION:
: APPLICANT: Wilson, David B.
: APPLICANT: Walker, Larry P.
: APPLICANT: Zhang, Sheng
: TITLE OF INVENTION: Thermostable Cellulase From A Thermomonospora Gene
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Hodgson, Russ, Andrews, Woods & Goodyear
: STREET: 1800 One Mt Plaza
: CITY: Buffalo
: STATE: New York
: COUNTRY: United States
: ZIP: 14203-2391
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
: OPERATING SYSTEM: MS-DOS/ Microsoft Windows 3.1
: SOFTWARE: Wordperfect for Windows 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/09069
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: U.S. Serial No. 08/265,429
: FILING DATE: 24/06/94
: ATTORNEY/AGENT INFORMATION:
: NAME: Nelson, M. Bud
: REGISTRATION NUMBER: 35,300
: REFERENCE/DOCKET NUMBER: 18617, 0008
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (716)856-4000
: TELEFAX: 716-849-0349
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1269 nucleotides
: TYPE: nucleic acid
: STRANDEDNESS: double-stranded
: TOPOLOGY: linear
: MOLECULE TYPE: genomic DNA
: ORIGINAL SOURCE:
: ORGANISM: Thermomonospora fusca
: STRAIN: YX36
: CELL TYPE: bacterium
: PCT-US95-09069-4

```

```

Query Match      8.5%; Score 314.6; DB 5; Length 1269;
Best Local Similarity 57.6%; Pred. No. 4,7e-52;
Matches 732; Conservative 0; Mismatches 479; Indels 60; Gaps 7;

QY 200 tegtcaaccctactctgtggcgaaagatctacagagcgagcgcgaaccacaatgcga 259
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 17 ttcacacaccccttgcaggggcccacagctgtacgtgacacccgctgtgctggcccaaggccg 76
QY 260 ctctgcagcgaataatgacgctgtgtttccacatattcgaagcgctgtgtatgtacgcga 319
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 77 cccttgagccggcgggttccgctgcgcgtgcacagatccacgctgtgtgctggaccgta 136
QY 320 tgcgtcgatcaacgagcgttcaacgagcgagcgcggttgaagataatctgaacgcgcgc 379
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 137 tggggcgcacatcgagggcaacagacagcccgacacacgagcctcattggtgtcgggacac 196
QY 380 t-----ctccagcagcaggaacacacccctgaatcatgtatgtatctacg 430
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 197 tggagagagccgttccgacagctccggttgccgacccgctgacacatccagctgtatctaca 256
QY 431 atctgcgagcagcagctgtgcggcgctgcctcccaacgagcagctgtccgcctacgag 490
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 257 acctgcccggccgcgactgcgcgcgctgtgccttccaaacgctgagctt-----ggtccg 310
QY 491 caggttgcagacctatgaacgcagatcacatccgattgcagatcctgtg---ca 547
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 311 atmaactcgacccgcttaccmaagcagatcattcaccgcatccgacatctatgtggagact 370
QY 548 atcgaagctactcgaagcgtgcgagatcgtgaacatcatgtacgcgagactcgtgcgaacg 607
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 371 tccacagactgcagaaactcgcgactcgtcccatcattcagatcgaactcctccgccaca 430
QY 608 cgtgcacaaatagagcattcaagcgtgtgtcaagcgtgtgcgtat----- 654
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 431 tcttcacacacggtggggggaacggccgacacagcttctccgcttaccatgaacacgaacg 490
QY 655 -----tacgagcagagcagatcagtagcgtctcgaagaatgcagccatccgaagctgt 709
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 491 gcggcttaccgtacacaggtgtgtggctacgcccctccgcaacgttgggcgagatccgaacgtct 550
QY 710 acactacatgtagcgcgcgccaactccgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 759
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 551 acaactacatcgacgacgcccacacacgctgagatcgactggactgacactccactccgcccct 610
QY 770 agctacaggaagtcggaaggtcctaag---cgagatcgggtgtcaagcgatcagc 826
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 611 cggctgacatcttctacagagccgacacacgcttccgcttccacgctgacgtgtgacg 670
QY 827 gcttcgtaaccaacagcggaattacacgctgtgaagagcgttcaatgaacgcgcaccc 886
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 671 gcttcatcttcaacacagccacacactacgctgacgctgtggagccgcttacttgagactcaacg 730
QY 887 agcaagctcggcgacagcgcggtgtgagtgcggaattctaccagtggaatctgtacatcg 946
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 731 gcacgcttaacggcagacgctatccgacagtcgaagtggtgactgaacacagatctgc 790
QY 947 acgaagccgactacgcggtgtgaattgtactcgcggtcgtgcgcgtgtgtgtgtgtgtgtgt 1006
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 791 acgagcttcccttgcacagagcctgtgacgcccctgacgcccgaacgcttccggtccg 850
QY 1007 gcatcgcatgtcatcgaacacttaacgaagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1066
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 851 acatcggatgctcatcgaacacttccgacacgctgtgggtgtggccacacgcttccgacg 910
QY 1067 gcccgagcagcgcagcagatgtcaaacacttgcacacagtcgaaagtgtgaacttcgtgc 1126
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 911 gaccgagctcttcacacgacactcaacacagcttgcagagccgctatgacgacgctgta 970
QY 1127 agcagcggcctgtgtgtgaacacgaaggtgcgggctcgcgcgacgcgcgcgaggaac 1186
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 971 tccaccccggttaactgtgtcaacacagccgctgctggcgtgacgacgccc----- 1023

```

QY 1187 gcccgaggaactcccgagacgcacactcgagcggtatgttgatcaagccgcgagtg 1246
Db 1024 -----ACGGTAAACCCGGCTCCCGGTGTGACGCTTACGTGTGGGTGAAGCCCGGGGTG 1078
QY 1247 agtcgagcgacacagcgctgcgagcgaltccgacaactcgcaagaagtcgagcccaatgt 1306
Db 1079 AGTCGAGCGGGCGCAGAGATCCCGAACGAGAGGCGCTTCGACCCGCAATGT 1138
QY 1307 ggcagccgagtcacagcagtcgtacggggtactgacccaaac-----gcgttacgga 1357
Db 1139 GCGACCCGACCTACAGGGCAACGCCGCAACGCGCAACCCCTCGGGGTGCGCTGCCCA 1198
QY 1358 acctccgacatcgccgagcgagtgctcccgagcgagttgacgagctgtcgcgagcgac 1417
Db 1199 ACGGCCCCATCTCCGGCCACTGTTCTTGCCCAAGTTCGCGAGCTGCGCAACGCGCT 1258
QY 1418 ggcagcgagtg 1428
Db 1259 ACCCGCCTCTG 1269

RESULT 6
US-09-136-574A-1
; Sequence 1, Application US/09136574A
; Patent No. 6294366
; GENERAL INFORMATION:
; APPLICANT: Farrington, Graham K.
; Anderson, Paige
; Gibbs, Moreland
; Bergquist, Peter
; Daniels, Roy
; Morgan, Hugh W.
; Williams, Diane P.
; TITLE OF INVENTION: Compositions and Methods for
; Treating Cellulose Containing Fabrics Using Truncated
; Cellulase Enzyme Compositions
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Center, P.O. Box 457
; CITY: Spring House
; STATE: PA
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/136,574A
; FILING DATE: 19-Aug-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/932,571
; FILING DATE: September 19, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: 1997US001/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11707 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-136-574A-1

Query Match 3.68; Score 134.2; DB 4; Length 11707;
Best Local Similarity 52.6%; Pred. No. 2.2e-17;
Matches 392; Conservative 0; Mismatches 338; Indels 15; Gaps 4;
QY 1559 gcccgagcccgagtcggcgagccgagtagtcgctcgctccgctccgagccgct 1618
Db 3835 CACCACACACCGGACACCGACCCCGACATGACGCGCACACCACTTCAACACCACTTG 3894
QY 1619 ctccgctcgccgagcccgagtcgagcccgagtagtcgctcgctccgagtcgga 1678
Db 3895 CACCGACATCAAGCCCGCACACCGACCGACAGTACAGCAAGCAAGCCGCTCAACGCCGA 3954
QY 1679 gccgctcccgctcgccgagcccgagtcgagcccgagtagtcgctcgctccgga 1738
Db 3955 GCGCCACAGTAGAGGTTACTGTGACTTCGACACACCAACCAACCGCGCGCACCGCA 4014
QY 1739 cgtcgccgctgctcg---gtggcgctgaaggttcagtlacaaagaatagatcgggc 1795
Db 4015 CAGGACACACTGGCACGGGGAAGTGTGAAAGTACTATACAAAGAACATGACCAAGTG 4074
QY 1796 cgggtgataacacagatacaaccgagtcctcagttgtgataacgagtcgctgctg 1855
Db 4075 CGAGCACAAGTTCTATTAAAGCCGCGTGTAAAGATGATGAGAGCAGCAAGTCTTG 4134
QY 1856 attgctgacggtgacggtgcggtacgtgtccaccggagatggtggtcgctgacactg 1915
Db 4135 ATCTTAGCAGGTTTAAGATAAGATCTGTACACAGTAGTAGTG---ACAAGCCACAGA 4191
QY 1916 tglataactgtgactgagcgcgagtggtgtggaatalccgccccttcgctcg 1975
Db 4192 GTGCGGTATGTGACTGGGCACAGATAGGGGCAAGCAATGTCAATTGTTGAAGC 4251
QY 1976 tgaacccgagcagcgcgagccgagacactacactgcaatgctgctgctgagac-- 2033
Db 4252 TGAGCAGCGGAGTAGTGAGGAGGATTTACTTGGAGGTAGATTTGACGTGAGCTG 4311
QY 2034 ----gttgccgctggtgtggtgcgagcggtgagatcaaaaccggtgataagatgact 2089
Db 4312 GCGAGTTCCAGCTGTGTAAAGACACAGGGGATATACAGATAGGTTTAAAGATGACT 4371
QY 2090 gtcgaattccacgagaccaatgactactcgatggagcaacacaccccttcagact 2149
Db 4372 GGAGCAATTACATACGACGACGACTGTGATGATGATGATGATGATGATGATGATG 4431
QY 2150 gga---cgaagtgacggtgacgtaacggtggtgtgtgtgtgtgtgtgtgtgtgtgt 2206
Db 4432 AGAATGCGAAGTAACTGCTGTATGATGATGATGATGATGATGATGATGATGATG 4491
QY 2207 gcaacgaccccgagcccgacacatcccgagcccgagcccgagcccgagcccgaggtg 2266
Db 4492 GAGCGACACTGTCACCGACAGCAAGCAACCAACGCAACTCCGACAGCAACGCA 4551
QY 2267 atgtgacgcgcgcgaggtgtgagc 2291
Db 4552 CACGACCGCCGACAGCAACGCCAAC 4576

RESULT 7
US-07-862-588B-1
; Sequence 1, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Ilnaa
; APPLICANT: Sch Ieln, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5916796disk of No. 5916796th America, Inc.
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.


```

; PRIOR APPLICATION NUMBER: PCT/US99/11971
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: 09/104,308
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 08/974,042
; PRIOR FILING DATE: 1997-11-19
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1116
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Nearest "neighbor" - Streptomyces
; OTHER INFORMATION: thermobaceous
US-09-321-981-2

Query Match      2.6%; Score 95.2; DB 4; Length 1116;
Best Local Similarity 51.2%; Pred. No. 4,6e-10;
Matches 381; Conservative 0; Mismatches 343; Indels 20; Gaps 6;

QY 2611 accaggtgcaaggcgacgaataacgggtgacagaccatgagtggaattcgtcgccca 2670
DB 121 accaccacgataccagaccggtactgtgtcagacacacgcgtggtgacagcgccacc 180
QY 2671 cagtgctcaccatacacaacgcgacggtgctctgacggtgagccactgcgaactcagc 2730
DB 181 cagtg-----cactcaatgtgaccggaacggttcgagatacccccagcgccgcttcg 234
QY 2731 ggttggaaccggtggtgcccgcgcgcgcgtatccgtcatcacaaagggtgacacgtgggc 2790
DB 235 gtgcgcacacacggtcgcccgacgaagtctcactcctcgctcagcagcgctgcacactagc 294
QY 2791 aactgcacacagaaagacgtcggtgacgcgcgattcagatcagatcgaattgttgctgtg 2850
DB 295 aactgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 351
QY 2851 acgtcgtgagtaagacgaaggtgtctgtcggtgcgtatagcgtggtcctacgcacatttg 2910
DB 352 agcagtgctctcctacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 411
QY 2911 acgaacagtatacccaagaacacgcgttcagcacaacggttaccgaataatgatcttgctg 2970
DB 412 ctggacaccacacacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 468
QY 2971 aattgcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3030
DB 469 aaccgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 525
QY 3031 ggtcacacgtggaatgtctgtgacaggttcagacagacgtcgtggaagatatttctactgc 3090
DB 526 ggcgcgacgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 582
QY 3091 ctgaaccctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3150
DB 583 ctgtgcctcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 642
QY 3151 ggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 3210
DB 643 gtacagcacaagcctgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 702
QY 3211 atctggcaaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3270
DB 703 cgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 761
QY 3271 agtctcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 3330
DB 762 -gggcgggaacggtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 820
QY 3331 tctcgaagcccgacccgttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 3354
DB 821 acacctgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 844
```

```

RESULT 11
US-09-197-649-7
; Sequence 7, Application US/09197649
; Patent No. 6194550
; GENERAL INFORMATION:
; APPLICANT: Gold, Larry
; APPLICANT: Tuerk, Craig
; APPLICANT: Prihnow, David
; APPLICANT: Smith, Jonathan D.
; TITLE OF INVENTION: Systematic Polypeptide Evolution by Reverse Translation
; FILE REFERENCE: NEX02/CI-CON
; CURRENT APPLICATION NUMBER: US/09/197,649
; EARLIER FILING DATE: 1998-11-23
; EARLIER APPLICATION NUMBER: 07/829,461
; EARLIER FILING DATE: 1992-01-31
; EARLIER APPLICATION NUMBER: 07/739,055
; EARLIER FILING DATE: 1991-08-01
; EARLIER APPLICATION NUMBER: 07/561,968
; EARLIER FILING DATE: 1990-08-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 7
; LENGTH: 390
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Sequence
; OTHER INFORMATION: having a 120 repeat of ACG flanked by fixed
; OTHER INFORMATION: fragments having NcoI restriction sites.
US-09-197-649-7

Query Match      2.6%; Score 94.8; DB 4; Length 390;
Best Local Similarity 55.0%; Pred. No. 4,8e-10;
Matches 186; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

QY 1418 ggcacagcgtgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1477
DB 2 ggcacagcgtgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 61
QY 1478 ctctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1537
DB 62 cgcacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 121
QY 1538 cgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1597
DB 122 cgcacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 181
QY 1598 cgtctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1657
DB 182 cgcacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 241
QY 1658 cgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1717
DB 242 cgcacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 301
QY 1718 gctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1755
DB 302 cgcacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 339

RESULT 12
US-09-197-649-7/c
; Sequence 7, Application US/09197649
; Patent No. 6194550
; GENERAL INFORMATION:
; APPLICANT: Gold, Larry
; APPLICANT: Tuerk, Craig
; APPLICANT: Prihnow, David
; APPLICANT: Smith, Jonathan D.
; TITLE OF INVENTION: Systematic Polypeptide Evolution by Reverse Translation
```

BEST PAGE AVAILABLE COPY

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☒ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

